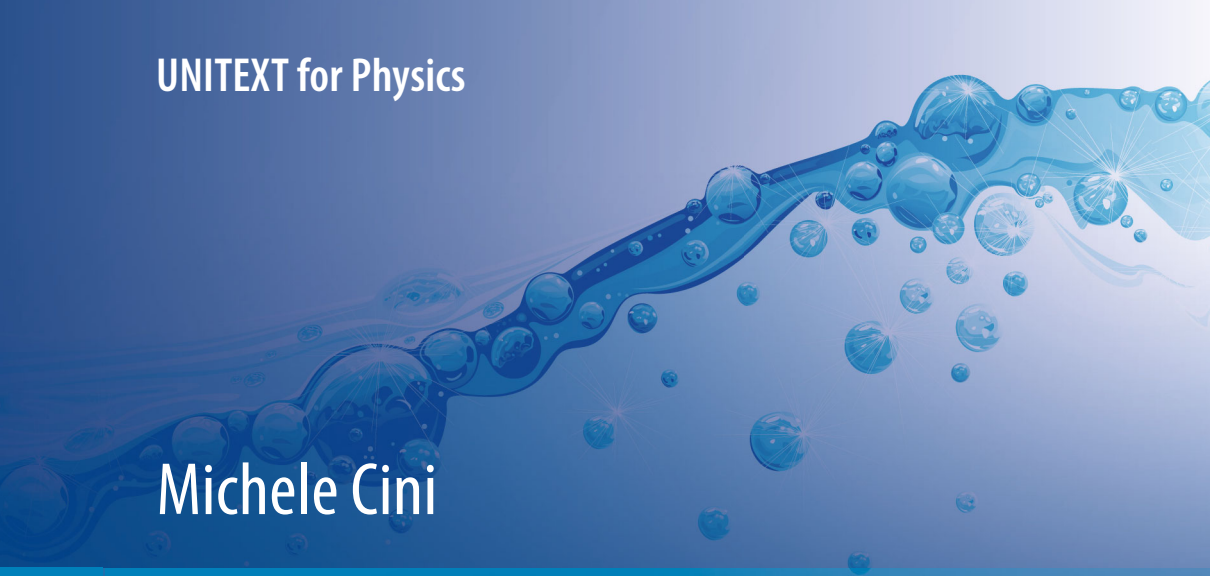
UNITEXT for Physics



Michele Cini

Elements of Classical and Quantum Physics

Second Edition

Springer

UNITEXT for Physics

Series Editors

Michele Cini, University of Rome Tor Vergata, Roma, Italy

Stefano Forte, University of Milan, Milan, Italy

Guido Montagna, University of Pavia, Pavia, Italy

Oreste Nicrosini, University of Pavia, Pavia, Italy

Luca Peliti, University of Napoli, Naples, Italy

Alberto Rotondi, Pavia, Italy

Paolo Biscari, Politecnico di Milano, Milan, Italy

Nicola Manini, University of Milan, Milan, Italy

Morten Hjorth-Jensen, Department of Physics and Astronomy, University of Oslo, Oslo, Norway

Alessandro De Angelis, Physics and Astronomy, INFN Sezione di Padova, Padova, Italy

UNITEXT for Physics series publishes textbooks in physics and astronomy, characterized by a didactic style and comprehensiveness. The books are addressed to upper-undergraduate and graduate students, but also to scientists and researchers as important resources for their education, knowledge, and teaching.

Michele Cini

Elements of Classical and Quantum Physics

Second Edition

Springer

Michele Cini
Università di Roma Tor Vergata
Rome, Italy

ISSN 2198-7882 ISSN 2198-7890 (electronic)
UNITEXT for Physics
ISBN 978-3-031-65440-4 ISBN 978-3-031-65441-1 (eBook)
https://doi.org/10.1007/978-3-031-65441-1

1st edition: © Springer International Publishing AG, part of Springer Nature 2018
2nd edition: © The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Preface

The second edition adds many new topics, but actually I have revised all the book. The chapters on Relativity have been largely rewritten, with special interest in black holes and other astrophysical applications of Einstein's General Theory and its developments. In the quantum part I have added several less well-known but intriguing developments. The choice was influenced (of course) by my background but also by the wish to focus on rapidly developing and promising fields in the simplest possible way. So the second edition includes an introductory treatment of Supersymmetry, the Kochen–Specker theorem, quantum optics, Graphene, the path integral with spin, Chern number, Auger line shapes and spontaneous symmetry breaking. There are new sections on superconductivity, the quantum Hall effect and topological insulators which have fundamental interest and are important for applications. Everywhere I have been driven by scientific curiosity and a desire to clarify new and interesting mathematical topics and techniques that are often ignored and/or misunderstood. In all the arguments, I have emphasized the boundary that divides Galilean science from far-fetched conjectures that tend to be presented as theories. I did not include Relativistic Quantum Mechanics and Relativistic Field Theories (beyond the quantization of the Electromagnetic Fields) on which there is already an easily available and excellent literature.

Rome, Italy
April 2024

Michele Cini

Contents

1 Theoretical Physics and Mathematics 1

2 Analytical Mechanics 5
2.1 Galileo's Revolution and Newton's $F = ma$ 5
2.2 Lagrangian Formalism 8
2.2.1 Rotating Platform 13
2.2.2 Kepler Problem 16
2.3 The Path Between t_1 and t_2 and the Action Integral 17
2.3.1 Principle of Least Action 18
2.4 Legendre Transformation 21
2.5 The Hamiltonian 22
2.5.1 Reduced Mass 24
2.5.2 Canonical Transformations 25
2.5.3 Hamilton–Jacobi Equation 29
2.5.4 Point Charge in an Electromagnetic Field 32
2.5.5 Poisson Brackets 35
2.6 Delaunay Elements 37
2.7 Noether Theorem 39
2.8 Chaos 41

3 Dirac's Delta 43
3.1 Definition of the δ 43
3.1.1 Volume of the Hypersphere in N Dimensions 49
3.1.2 Plancherel Theorem 50

4 Some Consequences of Maxwell's Equations 51
4.1 Fields and Potentials 51
4.2 Green's Function of the Wave Equation and Retarded Potentials 53
4.3 Lienard–Wiechert Potentials 56
4.4 Geometrical Optics and Fermat's Principle 57

4.5 Coherent Light . . . 59
4.5.1 The Measurement of Stellar Diameters: Hanbury Brown and Twiss Effect (Classical Case) . . . 59
4.6 The Hall Effect . . . 60
4.7 Sagnac Effect . . . 63

5 Thermal Physics . . . 65
5.1 The Principles of Thermodynamics . . . 65
5.2 Thermodynamic Potentials . . . 68
5.2.1 Equation of Clausius–Clapeyron . . . 70
5.2.2 Maxwell Relations . . . 70
5.2.3 Chemical Potential . . . 71
5.2.4 Perfect Classical Gas . . . 72
5.2.5 The Arrow of Time . . . 73
5.2.6 Low Temperatures and the Third Principle . . . 74
5.3 Black Body . . . 75
5.4 Classical Statistical Mechanics and Information . . . 80
5.5 Gibbs Averages . . . 81
5.6 Microcanonical Ensemble . . . 83
5.6.1 Entropy of the Perfect Gas . . . 86
5.7 Canonical Ensemble . . . 88
5.7.1 Canonical Temperature . . . 88
5.8 Canonical Distribution . . . 89
5.8.1 Hint to Shannon Entropy and Information Theory . . . 91
5.8.2 Maxwell Distribution of Molecular Energies . . . 92
5.8.3 Perfect Gas in the Canonical Ensemble and Boltzmann Statistics . . . 93
5.8.4 Theorem of Equipartition . . . 98
5.9 Ultraviolet Catastrophe and the Planck Formula . . . 99
5.10 Grand-Canonical Ensemble . . . 101
5.11 Random Numbers and Monte Carlo Methods . . . 102
5.11.1 Uses of Random Numbers . . . 102
5.11.2 Generation of Random Numbers . . . 103
5.11.3 Some Applications in Science . . . 103
5.12 Irreversible Gates and Toffoli Gate . . . 104
Reference . . . 105

6 Special Relativity . . . 107
6.1 Galileo's Ship . . . 107
6.2 From Thought Experiments to the Lorentz Transformation, Using Einstein's Undeclared Assumption . . . 109
6.2.1 Relativity of Simultaneity . . . 111
6.2.2 Relativity of Time . . . 112
6.2.3 Relativity of Lengths . . . 113
6.2.4 Relativity of Synchronization . . . 115
6.2.5 More Striking Paradoxes . . . 115

6.2.6 The Lorentz Transformation 116
6.2.7 Apparent Superluminal Motion and Other Spectacular Effects 117
6.2.8 Constant Acceleration, Rindler Metric and Interstellar Missions 118
6.2.9 Relativistic Addition of Velocities 120
6.3 The Geometry of Special Relativity: Minkowsky Chronotope 122
6.3.1 The Lorentz Transformation as a Rotation 124
6.3.2 Vectors and Tensors 125
6.3.3 Action of the Free Field 128
6.3.4 Doppler Effect and Aberration of Light 129
6.4 Relativistic Mechanics 133
6.4.1 Energy–Momentum Tensor (Alias Stress–Energy Tensor) 136
References 136

7 Curvilinear Coordinates and Curved Spaces 137
7.1 Calculus in Curved Spaces 141
7.1.1 Parallel Transport 141
7.1.2 Curvature 142
7.1.3 Christoffel Symbols 143
7.2 Geodesics 145
7.3 Curvature Tensor 147

8 Gravity 151
8.1 The Ehrenfest Paradox 151
8.2 Principle of Equivalence 152
8.3 The Principle of General Covariance and the Curved Space–Time 155
8.3.1 Space Geometry in Stationary Problems 156
8.3.2 Generalized Equation of Motion 160
8.3.3 Generalized Maxwell Equations 161
8.4 Einstein Field Equations 162
8.4.1 Weak Fields and Linearized Field Equations 164
8.5 Gravitational Collapse: True Facts, Conjectures and Myths 165
8.5.1 Schwarzschild Solution and Black Holes 165
8.5.2 Kruskal and Szekeres Coordinates 168
8.5.3 Eternal Black Holes, Daring Hypotheses and Wild Fantasies 170
8.5.4 Shift of the Perihelion 171

8.5.5 Geodetic Effect 174
8.5.6 Relativistic Delay of Signals in a Gravity Field 175
8.6 Clocks in a Gravity Field, GPS and Gravitational Redshift 176
8.7 Bending and Gravitational Lensing of Light 178
8.8 Kerr Black Hole, Frame Dragging and Gravitomagnetic Field 179
8.9 Gravitational Waves 182
8.9.1 Astrophysical Sources 183
8.9.2 Gravitons? 184
8.10 The Standard Model of Cosmology and the Big Bang 184
8.10.1 The Relativistic Models 185
8.10.2 Cosmic Expansion and Background Radiation 188
8.10.3 Dark Matter 189
8.10.4 Accelerated Expansion 190
8.11 Conjectures About Wormholes and Time Machines 192
8.12 A Fleeting Glimpse Into the Einstein–Schrödinger Unified Field Theory, Kaluza's Five-Dimensional Conjecture and the Brans–Dicke Theory 193
References 194

9 The Wave Function 197
9.1 Corpuscles and Waves in Classical Physics 199
9.2 Dualism 201
9.3 Which Way Did the Electron Pass Through? 203
9.3.1 The Bohm–Aharonov Effect 204
9.3.2 Experiments by Deutsch on Photons 205
9.4 Quantum Eraser Experiments and Delayed Choice Experiments 206
9.4.1 Plane Wave, Superposition Principle and Operators 208
9.5 The Collapse of the Wave Function and the Copenhagen Interpretation 211
Reference 214

10 The Schrödinger Equation and the Evolution Operator 215
10.1 The Schrödinger Equation 215
10.2 The Ehrenfest Theorem 218
10.3 The Stationary State Equation and Its Resolvent 218
10.3.1 Hard Potential Wall 219
10.3.2 Bound States and Continuum States 219
10.3.3 Resolvent 220
10.4 Continuity Equation 220
10.5 Schrödinger and Heisenberg Formulations of Quantum Mechanics 222

11 Particle in One Dimension 223
11.1 Deep Potential Well 223
11.2 Free Particle and Continuum States 225
11.2.1 Normalization 225
11.2.2 Phase Velocity 226
11.2.3 Group Velocity of a Wave Packet 227
11.3 Solutions of the Stationary State Equation in One Dimension 227
11.4 Potential Step 228
11.4.1 Energy Below Threshold 229
11.4.2 Energy Above Threshold 230
11.4.3 Potential Well of Finite Depth 231
11.4.4 Continuum States 233
11.5 The δ Potential 235
11.6 Potential Barrier and Tunneling 235

12 The Postulates of Quantum Mechanics: Postulate 1 241
12.1 The Wave Function 241
12.2 Gram–Schmidt Orthogonalization 242
12.3 Löwdin Symmetric Orthogonalization 242
12.4 Schwarz Inequality 244
12.5 Hilbert Spaces 244
12.6 Quantum Numbers 245

13 Postulate 2 247
13.1 Mathematical Context: Operators, Matrices and Function Spaces 247
13.2 Why Hermitian Operators? 248
13.3 Orthogonal Spaces 248
13.4 Completeness 249
13.5 Commutators: Angle and Angular Momentum 250
13.6 Commutators: Properties and Tricks 252
13.7 Angular Momentum in 3 Dimensions 253
13.7.1 The Algebra 253
13.7.2 Angular Momentum Matrices 254
13.8 Unitary Operators, Translations and Rotations 255
13.8.1 Translations 256
13.8.2 Rotations 256
13.9 General Uncertainty Principle and Compatible Observables 258
13.9.1 Canonical Transformations in Quantum Mechanics 261

14 Postulate 3 263
14.1 Reasons Why the Set Must Be Complete 263

15 Postulate 4 265
15.1 Time Derivative of an Observable and F = ma 265
15.1.1 Energy–Time Uncertainty 266
15.1.2 Adiabatic and Quasi-Adiabatic Evolution 267
15.1.3 Sudden Approximation 269
15.1.4 Connection to Statistical Mechanics 270
15.2 Supersymmetry in Quantum Mechanics 271
15.3 Galileo Transformation 273
15.4 Quantum Particle in Electromagnetic Field 274
15.4.1 The Local Phase Invariance Generates All of Electrodynamics! 275
15.5 Dirac's Magnetic Monopole 276
15.6 No Cloning Theorem 277
15.7 Path Integral Formulation of Quantum Mechanics, Part1: Particle Without Spin 278
15.7.1 Application of the Path Integral to the Theory of the Bohm–Aharonov Effect 281
15.8 Some Interpretations of Quantum Mechanics 283
15.8.1 Schrödinger's Cats 283
15.8.2 The De Broglie–Bohm Formulation of Quantum Theory 285
15.8.3 Measurement and Wave Function Collapse 285
References 287

16 The Quantum Harmonic Oscillator 289
16.1 From the Coordinate Representation to the Second Quantization Operators 290
16.2 Number Representation 293
16.2.1 Coherent States 296
16.3 Quantization of the Electromagnetic Field 297
16.3.1 Zero-Point Energy and Casimir Effect 299
16.3.2 Parametric Down Conversion and Squeezed Light 300

17 Stationary States of One Particle in 3 Dimensions 303
17.1 Separation of Variables in Cartesian Coordinates 303
17.2 Landau Levels 305
17.3 Separation of Variables in Spherical Coordinates 306
17.3.1 Spherical Harmonics 308
17.4 Central Field 309
17.5 The Hydrogenoid Atom 312
17.5.1 Coulomb Wave Functions 318
17.5.2 Atomic Shells 319
17.6 Lamb Shift and Radiative Corrections 319

17.7 Periodic Potentials, Floquet's Theorem and Bloch's Functions: Tight Binding Bands of Solids 321
17.7.1 Band Structure of Graphene 322
17.7.2 Su–Schrieffer–Heeger Model 324
Reference 325

18 Spin and Magnetic Field 327
18.1 Magnetic Moment and Angular Momentum in Classical Physics 327
18.2 The Stern–Gerlach Experiment 328
18.2.1 Spin of Elementary Particles and Bound States 331
18.3 Pauli Matrices 332
18.3.1 Euler-Angle Representation 335
18.3.2 How to Rotate a Spin 1/2? 335
18.3.3 Spin Magnetic Moment and Pauli Equation 337
18.4 Quantum "Sum" of Angular Momenta 339
18.4.1 Clebsch–Gordan Coefficients 341
18.5 Photon Spin and Quantum Cryptography 345
18.5.1 Rabi Model 346
18.6 Space–Time Symmetries of Pauli Electrons 347
18.7 Path Integral Formulation of Quantum Mechanics. Part 2: Spin 349
Reference 350

19 Systems of Particles and Second Quantization 351
19.1 The Pauli Principle 351
19.2 Two Identical Particles 353
19.2.1 Fermions 353
19.2.2 Bosons. Second Quantization 354
19.2.3 Helium 354
19.3 Many-Body Matrix Elements. Fermions 358
19.3.1 Second Quantization for Fermions 361
19.3.2 Fano–Anderson Models 366
19.4 Quasi-particles 367
19.4.1 Boson and Fermion Excitations 367
19.4.2 Anyons 368
19.5 Hellmann–Feynman Theorem and Virial Theorem 369
19.5.1 Hellmann–Feynman Theorem 369
19.5.2 Virial Theorem 369
19.6 Hint to Spontaneous Symmetry Breaking and Goldstone Theorem 370

20 Perturbation Theory 373
20.1 Time-Independent Rayleigh–Schrödinger Perturbation Theory 373
20.1.1 Discrete Non-degenerate Spectrum 374
20.1.2 Perturbation Theory for Degenerate Levels 378
20.2 Time-Dependent Perturbations 381
20.2.1 Periodic Perturbations and Fermi Golden Rule 383
20.3 The Evolution Operator 385
20.4 Heisenberg Representation and Finite Temperature 387
20.5 Interaction Representation 388
20.5.1 From Heisenberg Picture to Interaction Picture 389

21 Variational Principle for Schrödinger–Pauli Theory 391
21.1 The Ground State and the Absolute Minimum of Energy 391
21.2 Variational Approximations for One-Particle Problems 394
21.3 Hartree–Fock Method 395
21.4 The Density Functional (DF) Method 398
References 399

22 Discrete Models 401
22.1 Matrices and Useful Models 401
22.2 Gauge Changes in Lattice Models 402
22.3 Flux Quantization 403
22.4 Infinite Discrete Models and the Local Density of States 404
22.5 Exactly Solved Non-periodic Infinite Model 406
22.6 The Current Operator on a Discrete Basis 407

23 Pancharatnam Phase and Berry Phase 409
23.1 The Pancharatnam Phase 409
23.2 The Berry Phase 410
23.3 Chern Number 412
23.4 Molecular Aharonov–Bohm Effect 413
23.5 Su–Schrieffer–Heeger Model: Berry Connection and Winding Number 415
References 415

24 Fano Resonances 417
24.1 The Dynamics of the Excited States 417
24.2 Fano Model 419
24.3 Working in a Subspace: The Self-Energy Operator 421
24.4 Fano Line Shapes 423
24.5 Anderson Model and Hubbard Model 425
References 426

25 Quantum Statistical Mechanics 427
25.1 Temperature, Density Matrix and Decoherence Problems 427
25.2 Quantum Canonical Ensemble 428
25.3 Ideal Fermi Gas and Ideal Bose Gas 429
25.3.1 Bose–Einstein Statistics 431
25.4 Black Body and Photons 434
25.4.1 Einstein A and B Coefficients and the Principle of the LASER 436
25.4.2 Specific Heats of Solids 437
25.5 Fermi–Dirac Statistics 437
25.5.1 Fermi Gas 438
25.5.2 The Meitner–Auger Effect 442
25.5.3 Auger Line Shapes in Solids 444
25.5.4 Degeneracy Pressure 445
25.6 Conjectures on Quantum Gravity 446
25.6.1 Unruh Radiation 447
25.6.2 Black Hole Evaporation? 448
25.7 Hanbury Brown and Twiss Effect: Bosons and Fermions 450
References 451

26 Quantum Transport and Quantum Pumping 453
26.1 Fundamental Quanta of Resistance 453
26.1.1 Quantum Transport in Nanoscopic Devices 454
26.1.2 Magnetoresistance, Giant Magnetoresistance and Colossal Magnetoresistance 455
26.2 Characteristics and Transients 455
26.2.1 Landauer Formula 457
26.3 Quantum Pumping, Berry Phase and Chern Numbers 458
26.4 Hint to Spontaneous Symmetry Breaking and Goldstone Theorem 459
26.5 Superconductivity 461
26.5.1 Flux Quantization 462
26.5.2 Clues for Pairing 463
26.5.3 Superconductivity Theory: the Line of Reasoning 463
26.5.4 Cooper Pairs 464
26.5.5 The Bardeen–Cooper–Schrieffer (BCS) Hamiltonian 466
26.5.6 Why Perfect Diamagnetism and Zero Resistance? 471
26.5.7 The Phenomenological Richardson Model for Superconductivity 474
26.6 The Integer Quantum Hall Effect and Topological Insulators 475
26.6.1 Landau Levels in a Quasi Two-Dimensional System 476
26.6.2 The Staircase Paradox and the Revenge of the Surface 478

26.6.3 Topological Explanation of the Sharp Plateaux 481
26.6.4 Origin of the Oscillations of Longitudinal Resistivity 482
26.6.5 Hint to Topological Insulators 483
26.6.6 Very Short Hint to the Fractional Hall Effect 484
26.6.7 Spin Hall Effect 484
References 484

27 Entanglement, Its Challenges and Its Applications 485
27.1 What is the Electron Wave Function in the H Atom? 485
27.2 EPR Paradox 487
27.2.1 Bohm's Thought Experiment as a Formulation of the Paradox 488
27.2.2 Plot Twist: Bell's Inequality 488
27.2.3 Modified Born Thought Experiment for Testing Bell's Inequality 489
27.2.4 The Kochen–Specker Theorem and Quantum Contextuality 491
27.3 How the No-Cloning Theorem Saves Coexistence of Quantum Theory with Relativity? 492
27.4 Quantum Computer 493
27.5 Bell States and Quantum Teleportation 494
27.6 Weak measurements and Quantum Cheshire Cats 496
References 498

Index 499

Chapter 1
Theoretical Physics and Mathematics

Salvini:
It is true that the Copernican System perturbs Aristotle's Universe, but we are dealing with our Universe, the true, real one.
From Galileo Galilei, Dialogue Concerning the Two Chief World Systems.

Following the method of Galileo, Theoretical Physics uses Mathematics as natural and essential language to describe reality. But even Galileo would probably be surprised by the degree of success of the mathematical description of the world. Simple elegant laws like the Planck radiation formula agree with reality accurately over all the electromagnetic spectrum; on the other hand, the need for abstract mathematical concepts stemming from complex analysis, such as topology, Group Theory, and infinite-dimensional spaces shows that the fabric of the Universe is highly nontrivial. *Is God a mathematician?* is the title of an interesting book by Mario Livio. The Nobel Laureate in Physics Eugene Wigner in 1960 spoke about the *unreasonable effectiveness* of Mathematics in the understanding of Reality.[1] Like Mathematics, Physics involves a depth of thought that is unknown to most people; only well-posed questions are permissible, and there is no room for puns, since scientists deal with nothing but verifiable facts, which one cannot simply adjust to make them fit with one's statements. In fact, there is progress, and every scientist contributes to building something definitive, even if further progress may build on the new findings.

[1] This opens the question as to whether the findings of mathematicians must be considered as inventions like those of an artist or rather discoveries like those of a scientist. Personally I believe they are discoveries, since there is an internal logic in the development of Mathematics, and Mathematics is the logic of Nature.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_1

However, Theoretical Physics is not Mathematics. It has its own method of investigation and is subtle and concrete. While a Mathematician seeks the maximum of generality, the Theoretical Physicist must strive to give an operational meaning to all the new statements. Experiment has the final word on controversies. All the statements of a theorist make sense if they can be verified by experiment. This is not so, for example, in the case of recurrent theories about the Multiverse, which are often proposed but not of physical interest. Even the most abstract-looking aspects of the theory have an operational meaning, at least in principle. One must make an exception for Astrophysics, since a direct measurement of distances is not feasible and we rely on the known Laws of Physics and the consistency of clues.

The philosopher Carl Popper gained wide popularity by stressing that Science comes from experiment, so each of its conclusions is provisional, and at risk of being proved wrong the next morning by another experiment. This seems to remove Science as a source of a sound knowledge. But scientists have always known that, and have no orthodoxy to defend, and thus even try to test existing theories with ever-increasing precision measurements. Some discoveries have been completely unexpected and new surprises are possible. Relativity and Quantum Mechanics have shown that the extrapolation of physical laws to new ranges of phenomena is at risk. Besides, the borderline between Classical and Quantum Physics is still a matter of research. There are quantum phenomena, like entanglement, that have been tested over distances on the order of many kilometers in satellite experiments. Whatever new high energy physics may be discovered, the Maxwell equations and the laws of Thermodynamics are not in danger in regard to the wide range of phenomena for which they are known to work fine. Nevertheless, the 10 digits precision of the agreement between theory and experiment in Quantum Electrodynamics means that, besides the single experiment, there is a detailed and coherent understanding of many interdependent facts. The Truth exists in Religions. By far, a great number of the conclusions in Physics rank among the most sound knowledge that can exist. If somebody claims the invention of the perpetual motion, or superluminal flight, the claim can be safely ignored. The power of Science to increase the real grasp of Man and his ability to solve problems cannot be questioned.

Let us start with some constants (Table 1.1).[2] In the current Theories, some of the above values are really fundamental ($\hbar$, c, G, e) and enter as empirical constants. Nobody knows why the constants have such values, but a Universe with different values (if possible) should look very different. Others are known combinations of the former, while still others can be obtained through theories that are well established or in progress. For example, QCD calculations of the mass of the proton have been reported, in terms of quark masses, but such topics are outside the scope of the present book.

G was the first constant to be discovered (by Newton) and measured (by Cavendish, in 1798); Galileo tried to measure the speed of light with inadequate means; after the order-of-magnitude measurement by Rømer, we now know the much more accurate

[2] I am using both the International Units and the Gauss Units, since both are commonly used.

Table 1.1 Some constants of physics

Name	Symbol	Value	Units
Avogadro number	N_A	6,022169 10^{23}	mol^{-1}
Gravity constant	G	6,6732 10^{-11}	N m^2/Kg^2
Boltzmann constant	K_B	1,380622 10^{-23}	$J/^0K$
Stefan-Boltzmann constant	σ	5,66961 10^{-8}	$W/m^2 \times^0 K^4$
Speed of light	c	2,99792458 10^8	m/s
Classical radius of the electron	$r_e = \frac{e^2}{2mc^2}$	2,819489 10^{-15}	m
Electron mass	m_e	9,109558 10^{-31}	Kg
Proton mass	m_p	1,672614 10^{-27}	Kg
Neutron mass	m_n	1,674920 10^{-27}	Kg
Planck's constant	h	6,626196 10^{-34}	J s
Fine structure constant	$\alpha = \frac{e^2}{\hbar c}$	$\frac{1}{137,03602}$	Pure number, Gauss units
Fine structure constant	$\alpha = \frac{e^2}{2\epsilon_0 hc}$	$\frac{1}{137,03602}$	Pure number, SI units
Flux quantum	$\frac{hc}{e}$	4×10^{-7}	$Gausscm^2$
Electronvolt	eV	1,6 10^{-19}	J
Proton magnetic moment	μ_p	1,4106203 10^{-26}	J/T

value $c = 299792458$ m/s; the Maxwell equations, after a long crisis, led to Relativity. The introduction of the Boltzmann constant started Statistical Mechanics and led to a microscopic understanding of the Thermodynamic principles; the introduction of h by Planck started Quantum Mechanics.

The introduction of these constants has marked the great conceptual conquests that are sometimes called scientific revolutions. However, the term is too extreme.

It is true that Galileo tossed Aristotle's Physics into the basket, but Einstein's Relativity did not throw out Galileo, and Quantum Mechanics did not abolish Classical Physics.

The Mechanics of Galileo and Newton and its developments continue to be used in daily life; it describes well and correctly the motions of planets and stars; the Solar System is explored through the classical equations. Simply put, it cannot be extrapolated to high energies and phenomena that are far from the experimental conditions for which it was formulated. When c or h enter, we are outside the range of validity.

Notably, the generalized theories are not obtained by fixing details of the old ones, but require a new formalism and a set of new concepts; the older theories remain as limiting cases of the general theories. Understanding of the physical laws proceeds by successive layers. But we shall find that a deep part of the theory proceeds from the special to the general theory without any discontinuity.

Paradoxically, Science today has detractors and very active adversaries. But it remains the best result of man's ingenuity, our most important resource. The intuitions of the great scientists are unknown to most people but the effort which is needed to understand them is very rewarding.

Chapter 2
Analytical Mechanics

In the mathematical formulation of the classical theory, we meet key concepts that are needed in its extensions (Relativity, Quantum Mechanics). Theoretical Physics is a large body of knowledge, yet it is deeply unitary. No doubt, it begins here.

2.1 Galileo's Revolution and Newton's $F = ma$

Galileo Galilei[1] contradicted all the ancient wisdom and doctrine when he established the principle of inertia, clearly stated the principle of Relativity and measured the acceleration of gravity and the circular motion around the year 1600. His concept of Mechanics contrasted with that of Aristotle, who wrote that a force was needed to

[1] Galileo Galilei (Pisa 1564–Arcetri 1642) was the father of Modern Science. Professor of Mathematics in the University of Pisa from 1589, he moved to Padova and back to Pisa in 1611. He conceived and stated clearly for the first time the scientific method, based on experiments, thought experiments and mathematics. Actually, he was equally gifted for experiment, theory and publication of the results. He also tried to measure the speed of light and built the first astronomical telescope, becoming the founder of modern Astronomy. His *Sidereus Nuncius* (1610) is also a literary masterpiece for its clear and very readable style. When he reported the discovery of the *Pianeti Medicei* (Jupiter's moons) and of the sunspots, he was suspected of heresy. The foundations of Mechanics are in his *Discorsi e dimostrazioni matematiche intorno a due nuove scienze, attinenti alla meccanica ed ai meccanismi locali*, Leiden (1636). The book was published abroad because of the persecutions by the Church, as his teaching was contrary to any motion of the Earth. He avoided the stake, thanks to his prudence and good relations with the catholic hierarchy. He was a friend of Pope Urbano VIII Barberini. Nevertheless he was taken to trial, forced to retract and imprisoned for the rest of his life in Arcetri. Thereafter, life was very hard for scientists in Italy for quite a long time.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_2

keep a body in motion; and Aristotle's Physics had been given a holy status by the Counter-Reformation. Then, Newton[2] established the basic law of motion

$$m\overrightarrow{a} = \overrightarrow{F},$$

where $\overrightarrow{a}$ is the acceleration of a material point mass m under the action of a force $\overrightarrow{F}$. He also gave this *equation of motion* the well-known differential form. Newton's equation in one dimension reads:

$$m\frac{d^2x(t)}{dt^2} = F_x \tag{2.1}$$

that is, $m\ddot{x} = F_x$. In terms of the momentum $p_x = m\dot{x}$,

$$\dot{p_x} = F_x. \tag{2.2}$$

If F is known, this determines the *law of motion* $x(t)$ as a function of the initial conditions $x(0)$, $\dot{x}(0)$. The motion is then determined at all times. The laws of classical mechanics are these. The reader at this point might suspect that the rest of the chapter is just a series of examples of solution; instead there is a beautiful theory that we must build. The initial motivation for developing a theory was that the equation involves the second derivative and is not amenable to quadratures.

The *conservative* forces, which derive from a potential $V(x)$ according to

$$F_x = -\frac{dV(x)}{dx}, \tag{2.3}$$

are a promising special case with which to start an investigation, since they are easier. Along a trajectory $x(t)$, the potential $V(x(t))$ varies at a rate $\dot{V} = \frac{dV}{dx}\dot{x}$; then, multiplying (2.1) by $\dot{x}$, one finds $\dot{x}m\ddot{x} - \dot{x}F_x = \frac{dE}{dt}$, where E is the energy; this is the law of energy conservation. It implies that

$$E = T + V \tag{2.4}$$

is fixed. While $V(x(t))$ varies, the kinetic energy

$$T = \frac{m}{2}\dot{x}^2$$

[2] Isaac Newton (Woolsthorpe 1642–London 1727) was a professor in Cambridge since 1669; then he had already invented Calculus, although his book *De Methodis Serierum et Fluxionum* was published in 1671. The basic work about Mechanics is entitled *Philosophiae naturalis principia mathematica*, Iussu societatis regiae, London (1687). He subsequently discovered the law of gravity, and made discoveries in Optics. He also invented the calculus of variations and the reflector telescope. Unlike Galileo, he was not persecuted, but highly honored in his own country, receiving prestigious commissions from the Government.

also varies and the sum is constant. Equation (2.4) yields the *integral of the motion*

$$E = \frac{m}{2}\dot{x}^2 + V(x);$$

it links the first derivative $\dot{x}$ and the constant E, and is much simpler to solve than (2.1).

For a point mass in 3 dimensions, Newton's equation is a vector equation, and one writes $\overrightarrow{F}(\overrightarrow{r}) = -\overrightarrow{\nabla} V(\overrightarrow{r})$: for a system of N point masses, the equations are $3N$; interactions between particles produce coupled equations. Kinetic energies add:

$$T = \sum_{i}^{N} \sum_{\alpha}^{3} \frac{m_i}{2} \dot{x}_{i,\alpha}^2$$

(i runs over points and α over components of $\overrightarrow{r}_i$).

Analytical mechanics allows us to apply the laws of Newton effectively to any problem, even when the forces are not given at the start; above all, it reveals a mathematical structure underlying the equations of motion that is much more general than classical mechanics, as we shall see. The Theory of Relativity and Quantum Mechanics are generalizations of classical mechanics rooted in analytical mechanics.

In many interesting problems, the motion is limited by constraints. For example, a point particle may be constrained to move on a surface, or along a line. A constraint like this that can be expressed as an equation of the form $F(x_1, \ldots, x_N) = 0$ is called holonomic and reduces the number of unknowns and equations. For a point mass, one should solve 3 equations of motion, but if the particle is known to move on a circle of radius R, an angle is sufficient to locate the point. There is, however, the additional difficulty that the forces resulting from the constraints are known only after solving the problem.

For $N = 2$ independent points, one has 6 equations of motion to solve. Now suppose we have a system of two masses m_1, m_2 constrained to remain at a distance r_0 at the ends of a rigid stick of negligible mass, hinged at the origin of coordinates at the center of gravity. Such a system is called a *rigid rotor* in 3 dimensions. The center of gravity of the two masses is the point

$$\overrightarrow{R} = \frac{m_1 \overrightarrow{r}_1 + m_2 \overrightarrow{r_2}}{m_1 + m_2}. \tag{2.5}$$

In this case, $r_1 + r_2 = r_0$, and so $r_1 = \frac{m_2 r_0}{m_1 + m_2}$, $r_2 = \frac{m_1 r_0}{m_1 + m_2}$.

When the rotor moves, the stick acts on the masses with unknown forces. Are we lost? No. The general method for writing the equations of motion is due to Giuseppe Luigi Lagrange (Torino 1736–Paris 1813), the great Italian-French mathematician,

and (independently) to the great Swiss Euler.[3] Next, we shall see the Euler–Lagrange theory, assuming smooth constraints (that is, they should not make work and should not dissipate energy during the motion).

The rotor has two degrees of freedom, and we can choose the angles θ and ϕ of one mass as polar coordinates. We do not need 6 equations, but only 2. Before proceeding, I introduce the general method.

2.2 Lagrangian Formalism

Constraints and Lagrangian Coordinates

N material points in 3 dimensions are described in Cartesian coordinates by

$$\{x_1, y_1, z_1, x_2 \dots z_N\} \equiv \{x_{i\alpha}\}, \quad 0 < i \leq N, \quad 1 \leq \alpha \leq 3.$$

We consider constraints, such as those requiring the points to move on certain surfaces, or to maintain some fixed distances. The force on the particle i can be decomposed:

$$\overrightarrow{F}_i = \overrightarrow{F}_i^{appl} + \overrightarrow{F}_i^{vinc},$$

where the first term (applied force) is known (we can think of springs, gravitational forces, electric fields, assigned by the problem), but the constraint reaction is not. Suppose that the applied forces are conservative, i.e., they arise from a potential energy $V\{x_1, y_1, z_1, x_2 \dots z_N\}$ according to

$$F_{i,\alpha}^{appl} = -\frac{\partial V}{\partial x_{i,\alpha}}. \tag{2.6}$$

The $-$ sign tells us that the force pulls toward a decreased potential energy. We assume smooth constraints. Therefore, the reaction of the constraints is orthogonal to the infinitesimal shift of particle i $\delta \overrightarrow{R}_i$ and does no work:

$$\overrightarrow{F}_i^{vinc} \cdot \delta \overrightarrow{R}_i = 0. \tag{2.7}$$

The force due to the constraint is perpendicular to the shift of particle i. The work δL done by the applied forces is given by

[3] Leonhard Euler (Basel, Switzerland 1707–S. Petersburg, Russia 1783) was probably the greatest mathematician of all times. He contributed to all the fields of analysis, number theory and geometry. He also wrote the book *Mechanica* about his results on the equations of motion.

$$-\delta V = \delta L = \sum_i \vec{F}_i \cdot \delta \vec{R}_i \equiv \sum_i \vec{F}_i^{appl} \cdot \delta \vec{R}_i, \tag{2.8}$$

where i runs over the point masses; so

$$-\delta V = \sum_{i,\alpha} F_{i,\alpha}^{appl} \delta x_{i,\alpha}. \tag{2.9}$$

Note that we are assuming a potential that depends on positions and not on velocities. This restriction will be removed when dealing with the Lorentz force. Since the Cartesian shift $\{\delta x_1, \delta y_1, \delta z_1, \delta x_2 \ldots\} \equiv \{\delta x_{i\alpha}\}$ is not allowed by the constraints, the Lagrangian coordinates q_β, $1 \le \beta \le s$ are more convenient to represent $\delta \vec{R}_i$ provided that every δq_β represents a possible motion. One substantial advantage is that the number s of degrees of freedom is reduced: $s < 3N$. The Cartesian component of an infinitesimal motion allowed by the constraints is

$$\delta x_{i,\alpha} = \sum_\beta \frac{\partial x_{i,\alpha}}{\partial q_\beta} \delta q_\beta. \tag{2.10}$$

From (2.9), one gets

$$-\frac{\partial V}{\partial q_\beta} = \sum_{i,\alpha} F_{i,\alpha}^{appl} \frac{\partial x_{i,\alpha}}{\partial q_\beta}. \tag{2.11}$$

The infinitesimal work done by the applied force F^{appl} is

$$\delta V = \sum_\beta \frac{\partial V}{\partial q_\beta} \delta q_\beta. \tag{2.12}$$

In analogy with (2.6),

$$Q_\beta = -\frac{\partial V}{\partial q_\beta} \tag{2.13}$$

is called the generalized force; this can be written in terms of the applied force using (2.11):

$$Q_\beta = \sum_{i,\alpha} F_{i,\alpha}^{appl} \frac{\partial x_{i,\alpha}}{\partial q_\beta}. \tag{2.14}$$

The Euler–Lagrange Equations

Starting with the equations of motion and the constraints (if any) written in Cartesian coordinates, we want to rewrite everything in terms of Lagrangian coordinates q_α, $\alpha = 1, \ldots s$, in order to have as many coordinates as the degrees of freedom, and in order to take into account the constraints automatically. The scalar product of the Cartesian equation

$$\overrightarrow{F}_i = \frac{d\overrightarrow{p}_i}{dt} \tag{2.15}$$

with an infinitesimal displacement $\delta\overrightarrow{R}$ compatible with the constraints yields

$$-\delta V = \sum_i \overrightarrow{F}_i \cdot \delta\overrightarrow{R}_i = \sum_i \frac{d\overrightarrow{p}_i}{dt} \cdot \delta\overrightarrow{R}_i \Longrightarrow -\delta V = \sum_{i,\alpha} \dot{p}_{i,\alpha}\delta x_{i,\alpha}. \tag{2.16}$$

The first term is (2.12); the first equality introduces the assumption that the constraints do not make work. We change variables in the last term, using (2.10):

$$\sum_{i,\alpha} \dot{p}_{i,\alpha}\delta x_{i\alpha} = \sum_{i,\alpha} \dot{p}_{i,\alpha} \sum_\beta \frac{\partial x_{i\alpha}}{\partial q_\beta}\delta q_\beta = \sum_{i,\alpha} m_i\ddot{x}_{i,\alpha} \sum_\beta \frac{\partial x_{i\alpha}}{\partial q_\beta}\delta q_\beta;$$

exchanging the summations in order to highlight δq_β, one finds that

$$-\delta V = \sum_\beta \delta q_\beta \sum_{i,\alpha} m_i\ddot{x}_{i,\alpha}\frac{\partial x_{i\alpha}}{\partial q_\beta},$$

hence, we obtain the generalized force:

$$-\frac{\partial V}{\partial q_\beta} = \sum_{i,\alpha} m_i\ddot{x}_{i,\alpha}\frac{\partial x_{i\alpha}}{\partial q_\beta}.$$

The potential term was easy to rewrite in Lagrangian coordinates; to transform the kinetic term in Eq. 2.17, we note that

$$\ddot{x}_{i,\alpha}\frac{\partial x_{i\alpha}}{\partial q_\beta} = \frac{d}{dt}\left(\dot{x}_{i,\alpha}\frac{\partial x_{i\alpha}}{\partial q_\beta}\right) - \dot{x}_{i,\alpha}\frac{d}{dt}\frac{\partial x_{i,\alpha}}{\partial q_\beta},$$

that is, exchanging derivatives in the last term,

$$\ddot{x}_{i,\alpha}\frac{\partial x_{i\alpha}}{\partial q_\beta} = \frac{d}{dt}\left(\dot{x}_{i,\alpha}\frac{\partial x_{i\alpha}}{\partial q_\beta}\right) - \dot{x}_{i,\alpha}\frac{\partial \dot{x}_{i,\alpha}}{\partial q_\beta}. \tag{2.17}$$

So,

$$\sum_{i,\alpha} m_i\left[\frac{d}{dt}\left(\dot{x_{i\alpha}}\frac{\partial x_{ia}}{\partial q_\beta}\right) - \dot{x}_{i\alpha}\frac{\partial}{\partial q_\beta}\dot{x}_{i\alpha}\right] = -\frac{\partial V}{\partial q_\beta}. \tag{2.18}$$

The second term in the L.H.S. is

$$-\sum_{i,\alpha} m_i\dot{x}_{i\alpha}\frac{\partial}{\partial q_\beta}\dot{x}_{i\alpha} = -\frac{\partial T}{\partial q_\beta}, \tag{2.19}$$

where T is the kinetic energy. Now the equation of motion reads as:

$$\frac{d}{dt}\sum_{i,\alpha} m_i \left(\dot{x_{i\alpha}}\frac{\partial x_{ia}}{\partial q_\beta}\right) = -\frac{\partial}{\partial q_\beta}(T - V). \tag{2.20}$$

The Time Derivative

We wish to eliminate the Cartesian coordinates from the L.H.S., and this is highly nontrivial. First of all, dividing both sides of (2.10) by dt,

$$\dot{x}_{i\alpha} = \sum_\beta \frac{\partial x_{i\alpha}}{\partial q_\beta}\dot{q}_\beta. \tag{2.21}$$

At this point, the components of the velocity are to be thought of as functions of all the q and the time, but we decide that they also depend on the $\dot{q}_\beta$, which we take as independent variables. Differentiating $\dot{x}_{i\alpha}$ with respect to $\dot{q}_\beta$, one finds that

$$\frac{\partial \dot{x}_{i\alpha}}{\partial \dot{q}_\beta} = \frac{\partial x_{i\alpha}}{\partial q_\beta}. \tag{2.22}$$

Now the L.H.S. of (2.20) becomes

$$\frac{d}{dt}\sum_{i,\alpha} m_i \left(\dot{x_{i\alpha}}\frac{\partial \dot{x_{ia}}}{\partial \dot{q}_\beta}\right) = \frac{d}{dt}\frac{\partial T}{\partial \dot{q}_\beta}. \tag{2.23}$$

Substitution in (2.20) yields

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{q}_\beta} = \frac{\partial T}{\partial q_\beta} - \frac{\partial V}{\partial q_\beta}. \tag{2.24}$$

Then, we introduce the Lagrangian function, or simply the Lagrangian

$$L(q, \dot{q}, t) = T(q, \dot{q}, t) - V(q, t), \tag{2.25}$$

where q stands for the set of all coordinates, $\dot{q}$ for the set of velocities and so on. The Lagrangian equations of motion, valid in every reference and coordinate system, are of the same form for all systems, namely

$$\boxed{\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_\beta} = \frac{\partial L}{\partial q_\beta}.} \tag{2.26}$$

To appreciate the reasons why $L = L(q, \dot{q}, t)$, we must pause for a moment to understand the decisive step taken in Eq. (2.22). The fact that the velocity components are to be treated as independent variables in that particular way is not a physically

evident fact, but it is the nontrivial point, the one that required the work of great mathematicians. Indeed, $\dot{q}_\beta$ is determined if we know q_β versus time t, so one could have taken $\dot{q}$ as a function of q and t. This would have led to a mathematically different functional dependence and an impasse. The Lagrangian formalism is achieved only by considering the velocities as independent variables for T and L.

The *generalized momenta* are defined by

$$p_i = \frac{\partial L}{\partial \dot{q}_i} \tag{2.27}$$

and are functions of the independent variables of L, namely

$$p_i = p_i(q, \dot{q}, t), \tag{2.28}$$

(recall that, here, $q, \dot{q}$ stand for the set of all the coordinates and velocities). If, for a particular q, $\frac{\partial L}{\partial q} = 0$, then q is *cyclic* and Eq. (2.26) says that the corresponding momentum (which is called the conjugate momentum) is conserved. If an angle is cyclic, a component of the angular momentum is conserved; if x is cyclic, p_x is conserved and so on.

If L has no explicit dependence on time, the energy is conserved. Indeed, the total derivative is

$$\frac{dL}{dt} = \sum_\alpha \frac{\partial L}{\partial q_\alpha}\dot{q}_\alpha + \sum_\alpha \frac{\partial L}{\partial \dot{q}_\alpha}\ddot{q}_\alpha, \tag{2.29}$$

and one finds that

$$\frac{dL}{dt} = \sum_\alpha \dot{q}_\alpha \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_\beta} + \sum_\alpha \frac{\partial L}{\partial \dot{q}_\alpha}\ddot{q}_\alpha = \frac{d}{dt}\sum_\alpha \frac{\partial L}{\partial \dot{q}_\alpha}\dot{q}_\alpha,$$

which implies the conservation of the energy

$$\sum_\alpha \frac{\partial L}{\partial \dot{q}_\alpha}\dot{q}_\alpha - L = \sum_\alpha p_\alpha \dot{q}_\alpha - L = E. \tag{2.30}$$

We have a great freedom of choice of Lagrangian coordinates q, and the equations of motion are always correct. In other words, the theory is invariant under *point transformations* to new coordinates $Q = Q(q, t)$. These transformations include the change of reference to an arbitrarily moving frame. If one knows a Lagrangian $L(q, \dot{q}, t)$ and wants to switch to a new Lagrangian $L(Q, \dot{Q}, t)$, all that is needed is to express the old variables in terms of the new, i.e., to make a change of variables. The reason for this statement will be clear when we introduce the action, below.

2.2.1 Rotating Platform

Let us write the equations of motion of a material point in a platform rotating in a clockwise direction with frequency ν. Since the Earth is a rotating frame, this example is relevant to everyday life. Let $\omega = 2\pi\nu$. In the inertial frame, the Lagrangian reads

$$L(x_0, y_0, \dot{x}_0, \dot{y}_0) = \frac{1}{2}m(\dot{x}_0^2 + \dot{y}_0^2), \tag{2.31}$$

and the point transformation is of the form

$$x = x_0\cos(\omega t) - y_0\sin(\omega t),\ y = y_0\cos(\omega t) + x_0\sin(\omega t). \tag{2.32}$$

We find that the square of the velocity is $\dot{x}_0^2 + \dot{y}_0^2 = (\dot{x} + \omega y)^2 + (\dot{y} - \omega x)^2$. Thus, the Lagrangian in the rotating frame becomes

$$L(x, y, \dot{x}, \dot{y}) = \frac{1}{2}m[\dot{x}^2 + \dot{y}^2 + \omega^2(x^2 + y^2) - 2\omega(x\dot{y} - y\dot{x})]. \tag{2.33}$$

The equations of motion are as follows:

$$m\ddot{x} = m\omega^2 x - 2\,m\omega\dot{y}, \qquad m\ddot{y} = m\omega^2 y + 2\,m\omega\dot{x}. \tag{2.34}$$

In the rotating system, we experience the obvious centrifugal force

$$\overrightarrow{F} = (F_x, F_y) = m\omega^2(x, y) \tag{2.35}$$

and the less obvious velocity-dependent Coriolis force

$$2m\omega(-\dot{y}, \dot{x}). \tag{2.36}$$

Problem 2.1 Write the Lagrangian and the equations of motion for the one-dimensional oscillator of mass m and force constant k, that is, $F = -kx$.

Solution 2.1

$$L = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2.$$

Problem 2.2 A plane pendulum has mass m and length l; the acceleration of gravity which is parallel to the vertical axis z is g. Write the Lagrangian and the equation of motion.

Solution 2.2 The speed is $l\dot{\phi}$, and the kinetic energy is $T = \frac{1}{2}ml^2\dot{\phi}^2$. Since $V = mgz$, one finds $L = \frac{1}{2}ml^2\dot{\phi}^2 + mgl\cos(\phi)$. Therefore, $\frac{\partial L}{\partial \dot{\phi}} = ml^2\dot{\phi}$, $\frac{\partial L}{\partial \phi} = -mgl\sin(\phi)$ and the equation of motion is $l\ddot{\phi} = -g\sin(\phi)$.

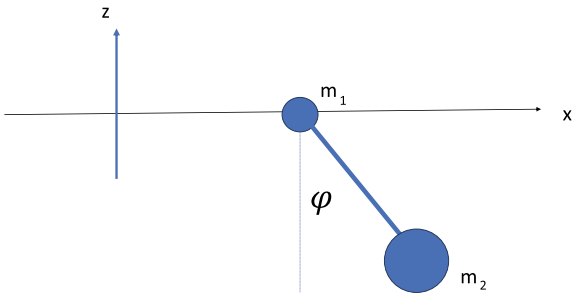


Fig. 2.1 A pendulum hinged to a mobile mass

Problem 2.3 A plane pendulum of mass m_2 and length l is linked to a mass m_1, which can move on a horizontal axis x. The acceleration of gravity is parallel to the vertical axis z and its magnitude is g. Write the Lagrangian.

Solution 2.3 Since $x_2 = x + l\sin(\phi)$, $z_2 = -l\cos(\phi)$, $\dot{x}_2 = \dot{x} + l\cos(\phi)\dot{\phi}$, $\dot{z}_2 = l\sin(\phi)\dot{\phi}$, and substituting into $T = \frac{1}{2}[m_1\dot{x}^2 + m_2(\dot{x}_2^2 + \dot{z}_2^2)]$ one finds $T = \frac{1}{2}(m_1 + m_2)\dot{x}^2 + \frac{m_2}{2}(l^2\dot{\phi}^2 + 2l\dot{x}\dot{\phi}\cos(\phi))$. So,

$$L = \frac{1}{2}(m_1 + m_2)\dot{x}^2 + \frac{m_2}{2}(l^2\dot{\phi}^2 + 2l\dot{x}\dot{\phi}\cos(\phi)) + m_2 gl\cos(\phi).$$

Hence, the equations of motion follow immediately (Fig. 2.1).

Problem 2.4 (*Free-falling oscillator*) In a free-falling lift (the acceleration of gravity is g), a harmonic oscillator (mass $= m$, elastic constant $= k$, force $F = -kz$) is aligned to the vertical z axis of a local Cartesian system (x, y, z), which is fixed in the lift. The oscillator starts from $z = 0$ at time $t = 0$. At $t = 0$, the local system (x, y, z) coincides with a stationary one (XYZ) fixed in the ground and the relative velocity is 0. Write the Lagrangian, equation of motion and its general solution in the (XYZ) frame.

Solution 2.4 The Lagrangian in the system of the lift is $L = \frac{1}{2}m\dot{z}^2 - \frac{1}{2}kz^2$. The origin of the falling system is at $Z = -\frac{1}{2}gt^2$ and the oscillator is at $Z(t) = z(t) - \frac{1}{2}gt^2$. Changing the coordinates,

$$L = \frac{1}{2}m\dot{z}^2 - \frac{1}{2}kz^2 = \frac{1}{2}m\left(\dot{Z} + gt\right)^2 - \frac{1}{2}k\left(Z + \frac{1}{2}gt^2\right)^2,$$

the equation of motion reads as:

$$m(\ddot{Z} + g) = -k\left(Z + \frac{1}{2}gt^2\right).$$

The general solution is $Z(t) = A\sin(\omega t + \varphi) - \frac{1}{2}gt^2$.

Problem 2.5 Rigid rotor. Above, we have defined the rotor as a system of two masses m_1, m_2 constrained to remain at a distance r_0 at the ends of a rigid stick of negligible mass. Find the equations of motion for the rotor.

Solution 2.5 The coordinate transformation is

$$\begin{aligned} x_1 &= r_1 \ \sin\theta \ \cos\phi, \\ y_1 &= r_1 \ \sin\theta \ \sin\phi, \\ z_1 &= r_1 \ \cos\theta, \end{aligned}$$

with $x_2 = -\frac{r_2}{r_1}x_1 = -\frac{m_1}{m_2}x_1$, and so on. The Lagrangian coincides with the kinetic energy, that is, in obvious notation,

$$T = \frac{1}{2}(m_1 \overrightarrow{v}_1^{\,2} + m_2 \overrightarrow{v}_2^{\,2}).$$

From the velocity components

$$\begin{aligned} \dot{x}_1 &= r_1(\cos\theta\cos\phi\dot{\theta} - \sin\theta\sin\phi\ \dot{\phi}), \\ \dot{y}_1 &= r_1\ (\cos\theta\sin\phi\dot{\theta} + \sin\theta\ \cos\phi\dot{\phi}), \\ \dot{z}_1 &= -r_1\ \sin\theta\ \dot{\theta}, \end{aligned}$$

one finds $\overrightarrow{v}_1^{\,2} = \dot{x}_1^2 + \dot{y}_1^2 + \dot{z}_1^2 = r_1^2(\dot{\theta}^2 + \sin^2\theta\dot{\phi}^2)$, $\overrightarrow{v}_2^{\,2} = r_2^2(\dot{\theta}^2 + \sin^2\theta\dot{\phi}^2)$. Therefore,

$$T = \frac{1}{2}I(\dot{\theta}^2 + \sin^2\theta\dot{\phi}^2),$$

where

$$I = m_1 r_1^2 + m_2 r_2^2$$

is the *moment of inertia* of the rotor. Setting

$$\begin{aligned} p_\theta &= \frac{\partial L}{\partial\dot{\theta}} = I\dot{\theta}, \\ p_\phi &= \frac{\partial L}{\partial\dot{\phi}} = I\sin^2\theta\dot{\phi}, \end{aligned}$$

we may write the equation of motion

$$\begin{aligned} \dot{p}_\theta &= I sin\theta\cos\theta\dot{\phi}^2, \\ \dot{p}_\phi &= 0. \end{aligned}$$

Problem 2.6 A point mass m moves along the curve $z(x) = L\sin(\frac{x}{L})$ on the vertical zx plane, where L is some length. There is no friction and the acceleration of gravity is g. One can take x as the only Lagrangian coordinate of this one-dimensional motion. Write the Lagrangian $L(x, \dot{x})$ and the kinetic momentum p_x.

Solution 2.6 $\dot{z} = \cos(\frac{x}{L})\dot{x}$, therefore setting $c(x) \equiv \cos\left(\frac{x}{L}\right)$,

$$T = \frac{1}{2}m\dot{x}^2\left[1 + c^2(x)\right], L = T - mgL\sin\left(\frac{x}{L}\right).$$

Hence, $p_x = \frac{\partial L}{\partial \dot{x}} = m\dot{x}\left[1 + c^2\right]$.

2.2.2 Kepler Problem

The two-body problem, in which the two bodies interact through a central force F that varies as the inverse square of the distance r between them, is known as the Kepler problem. In Sect. 2.5.1, we shall see that a two-body problem of this kind can be reduced to the problem of a single effective mass in an external potential. Here, we review the latter problem, first solved by I. Newton.

In cylindric coordinates, the square length element is $dl^2 = dr^2 + r^2 d\phi^2 + dz^2$, and therefore the kinetic energy of a point mass m is $K = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\phi}^2 + \dot{z}^2)$. Letting the orbit of a planet of mass m lay on the z plane, the Lagrangian is, in obvious notation,

$$L = \frac{1}{2}m\left[\dot{r}^2 + r^2\dot{\phi}^2\right] + G\frac{Mm}{r}. \tag{2.37}$$

One readily obtains from the Lagrange formalism the conservation of the angular momentum directed along z $J = \frac{\partial L}{\partial \dot{\phi}} = mr^2\dot{\phi}$ and the energy $E = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\phi}^2) - \frac{GMm}{r}$. These conditions determine the orbits, but we need to do some work to extract them. We divide both sides by

$$\dot{\phi}^2 = \left(\frac{J}{mr^2}\right)^2 \tag{2.38}$$

and make the substitution

$$\frac{\frac{dr}{dt}}{\frac{d\phi}{dt}} \to \frac{dr}{d\phi};$$

we arrive at

$$\frac{mr^4E}{J^2} = \frac{m^2r^4}{2J^2}\left(\frac{dr}{d\phi}\frac{d\phi}{dt}\right)^2 + \frac{m^2r^6}{2J^2}\left(\frac{d\phi}{dt}\right)^2 - \frac{Gm^2Mr^3}{J^2}. \tag{2.39}$$

Next, let $\xi = \frac{1}{r}$, which implies $\frac{dr}{d\phi} = -\frac{1}{\xi^2}\frac{d\xi}{d\phi}$, and use Eq. (2.38); this gives us

$$\frac{mE}{J^2} = \frac{1}{2}\left(\frac{d\xi}{d\phi}\right)^2 + \frac{\xi^2}{2} - \frac{GMm^2\xi}{J^2}. \tag{2.40}$$

Next, we apply $\frac{\partial}{\partial\phi}$ and simplify, obtaining

$$\frac{d^2\xi}{d\phi^2} + \xi - \frac{GMm^2}{J^2} = 0. \tag{2.41}$$

This leads us to

$$\xi = \frac{GMm^2}{J^2}[1 + e\cos(\phi - \phi_0)]. \tag{2.42}$$

Then, the orbit is given by

$$r = \frac{a(1-e^2)}{1 + e\cos(\phi - \phi_0)}, \tag{2.43}$$

with $e = \sqrt{1 + \frac{2EJ^2}{G^2M^2m^3}}$, $a(1-e^2) = \frac{J^2}{GMm^2}$. If the constant of integration e is less than 1, the orbit is an ellipse with eccentricity e and major axis $2a$. If $e > 1$ it is a hyperbola, and for $e = 1$ it is a parabola.

2.3 The Path Between t_1 and t_2 and the Action Integral

The Lagrangian contains the necessary information for writing the equation of motion, but it is not a measurable quantity, and every change of coordinates changes it as a function of time. Is there any quantity that can be used to characterize a given motion and is invariant for punctual transformations? To answer this question, we reformulate the problem of Mechanics as follows. Suppose we know the forces that act in a mechanical system, and we observe that in the time interval (t_1, t_2), the coordinate q goes[4] from $q(t_1)$ to $q(t_2)$. In a sense we have two *snapshots*. Can we use this information to retrace what happened in between? The answer must be sought with the help of the Lagrangian. A priori, i.e., before solving the problem of motion, one knows only that the physical law $q(t)$ belongs to a large set $\mathcal{E}$ of functions $q(t)$ which have nothing in common except that they all satisfy the same boundary conditions; these trajectory laws are called *virtual paths*. Solving the equations of motion means choosing between the virtual paths $q(t) \in \mathcal{E}$ the few possible ones. For every

[4] As usual, we speak about one variable q for short, but the argument is intended for a set of s variables.

virtual path $q(t)$, it turns out that $L(q(t), \dot{q}(t), t)$ is a function of time. Integrating this function of time, we can calculate the *action* S:

$$S = \int_{t_1}^{t_2} dt\, L(q(t), \dot{q}(t), t). \tag{2.44}$$

Dimensionally, S is energy $\times$ time and is a *functional* of the virtual path, i.e., a function of all the values of $q(t)$ at each t in the path under consideration. In this way, S depends on an infinite number of variables. For instance, for a point mass moving at constant speed $v = \frac{x_b - x_a}{t_2 - t_1}$ from x_a to x_b,

$$S = \frac{m}{2} \int_{t_1}^{t_2} dt \frac{(x_b - x_a)^2}{(t_2 - t_1)^2} = \frac{m}{2} \frac{(x_b - x_a)^2}{t_2 - t_1}. \tag{2.45}$$

There is an infinity of alternative virtual paths $x(t)$, $x(t_1) = x_a$, $x(t_2) = x_b$, and one can calculate S for each of them.

The reason why S is interesting is that it does not depend on the choice of Lagrangian coordinates used to describe the virtual path, but only on the path itself. In fact, the value that S takes at time t does not change if we make a point transformation, which is simply a change of variables in an integral. The fact that, for a given path, S is not arbitrary suggests that it must have a physical meaning, and, indeed, we shall see that it offers the opportunity to make considerable further progress.

Problem 2.7 For a harmonic oscillator of mass m and constant k, $L = \frac{m}{2}(\dot{x}^2 - \omega^2 x^2)$, where $\omega^2 = \frac{k}{m}$. Calculate S for a physical path from $x = 0, t = 0$ to $x = X, t = T$.

Solution 2.7 Set $x = A \sin(\omega t)$, $X = A \sin(\omega T)$, $L = \frac{m}{2} A^2 \omega^2 \cos(2\omega t)$; one finds

$$S = \frac{m}{2} A^2 \omega \sin(\omega T) \cos(\omega T) = \frac{m}{2} X^2 \omega \mathrm{cotg}(\omega T).$$

2.3.1 Principle of Least Action

There is an infinity of virtual paths between the *snapshots* $q(t_1)$ and $q(t_2)$, and one should distinguish those that are physically possible (if any) from the infinity of absurd ones. The action integral assigns a number (with the dimensions of energy $\times$ time, or angular momentum) to each virtual path, and so we can compare the action integrals of different paths with the initial and final configurations $q(t_1)$ and $q(t_2)$. Two virtual paths that differ little in position and velocity also give close results when one computes S. In the infinite-dimensional space of paths, one has the means to compare two of them and also define the concept of the extreme, such as a maximum or a local minimum. For instance, a path is a minimum if any "small" change produces an increase of S. In order to define rigorously an arbitrary variation,

we define an arbitrary function $\eta_i(t)$ such that $\eta_i(t_1) = \eta_i(t_2) = 0$ and vary the path by $q_i(t) \to q_i(t) + \eta_i(t)$. The varied Lagrangian is

$$L(q_i(t), \dot{q}_i(t), t) \to L(q_i(t) + \eta_i(t), \dot{q}_i(t) + \dot{\eta}_i(t), t).$$

We are interested in small arbitrary variations, therefore we instead put

$$q_i(t) \to q_i(t) + \alpha_i \eta_i(t),$$

where α_i is a parameter that allows us to adjust the magnitude of the change. We can always (provided that $\eta_i(t)$ is not too bad) choose α_i small enough to obtain a small change of S. For a generic $q(t)$, with $\alpha_i \ll 1$ we find, when we vary paths,

$$S(\alpha_i) - S(0) = \delta S^{(1)} \alpha_i + \delta S^{(2)} \alpha_i^2 + \ldots;$$

and the variational principle is

$$\delta S^{(1)} = \frac{\partial S}{\partial \alpha_i} d\alpha_i = 0. \tag{2.46}$$

This condition makes the action stationary; the first-order correction vanishes.

Since

$$S(\alpha_i) = \int_{t_1}^{t_2} dt L(q_i(t) + \alpha_i \eta_i(t), \dot{q}_i(t) + \alpha_i \dot{\eta}_i(t), t) dt,$$

the condition is

$$\frac{\partial S}{\partial \alpha_i}|_{\alpha_i \to 0} = 0 = \int_{t_1}^{t_2} dt \left[\frac{\partial L}{\partial q_i} \eta_i(t) + \frac{\partial L}{\partial \dot{q}_i} \dot{\eta}_i(t) \right]. \tag{2.47}$$

We must transform $\dot{\eta}$, since we do not know its behavior at t_1 and t_2, so we integrate by parts:

$$\frac{\partial S}{\partial \alpha_i} = 0 = \left[\frac{\partial L}{\partial \dot{q}_i} \eta_i(t) \right]_{t_1}^{t_2} + \int_{t_1}^{t_2} dt \left[\frac{\partial L}{\partial q_i} \eta_i(t) - \eta_i(t) \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} \right]. \tag{2.48}$$

Since $\eta_i(t)$ vanishes at t_1 and t_2,

$$\int_{t_1}^{t_2} dt \eta_i(t) \left[\frac{\partial L}{\partial q_i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} \right] = 0. \tag{2.49}$$

This is possible with arbitrary η_i only if the Euler–Lagrange equations hold. Therefore, the principle of least action is a particularly concise way to reformulate the laws of mechanics.

The fascination for variational principles has always been strong. Pierre Louis de Maupertuis, who wrote about this principle in 1744, felt that "Nature is thrifty in all its actions." Earlier, Pierre de Fermat (1601–1665) found a famous variational principle in optics, which is presented in Sect. 4.4. In Theoretical Physics, variational principles are everywhere. They are important in Electromagnetism, General Relativity and Quantum Mechanics. Besides a great aesthetic *appeal,* they are also important in practice, as we shall see.

It is not granted that the least action principle leads to a unique extremal path. It is not always possible to retrace what happened in between two *snapshots* giving the configuration of a system at different times, i.e., $q(t_1)$ and $q(t_2)$. If the two snapshots show a pendulum in the vertical position, it could have always remained still, or it could have been in oscillation provided that $t_2 - t_1$ is a multiple of the period of oscillation. In general, this sort of information selects a class of possibilities, but the motions that are not compatible with the Euler–Lagrange equations are impossible.[5] One learns from this example that the physical motions correspond to stationary paths but not necessarily to the minimum.

Suppose we add the total time derivative of an arbitrary function to the Lagrangian:

$$L(q, \dot{q}, t) \to L(q, \dot{q}, t) + \frac{d}{dt} F(q, t). \tag{2.50}$$

This addition changes S by $F(q(t_2), t_2) - F(q(t_1), t_1)$. The equations of motion are obtained by varying the paths; however, the initial and final "snapshots" are not changed. So, one is free to add the total time derivative of an arbitrary function to the Lagrangian, since this does not alter δS. The Lagrangian is largely arbitrary, and one can exploit this fact.

While $\dot{q}$ are independent variables for L, their variation is a consequence of the variation of q. If $q_i(t, \alpha_i) = q_i(t) + \alpha_i \eta_i(t)$, then $\dot{q}_i(t, \alpha_i) = \dot{q}_i(t) + \alpha_i \dot{\eta}_i(t)$. So, there is some redundancy in the process.

The variational principle contains another remarkable result, if instead of the virtual paths from $q_1(t_1)$ to $q_2(t_2,)$ we consider the physical paths, without fixing $q_2(t_2)$. Then, δS does not vanish any more, but instead of Eq. 2.48 one finds

$$\frac{\partial S}{\partial \alpha_i} = \left[\frac{\partial L}{\partial \dot{q}_i} \eta_i(t) \right]_{t_1}^{t_2} + \int_{t_1}^{t_2} dt \left[\frac{\partial L}{\partial q_i} \eta_i(t) - \eta_i(t) \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} \right], \tag{2.51}$$

where the integral now vanishes; since $\eta_i(t_2) = \frac{\partial q_i}{\partial \alpha_1}|_0$, we may write $\delta S = \sum_i \frac{\partial L}{\partial \dot{q}_i} \delta q_i(t_2)$, that is,

$$\frac{\partial S}{\partial q_i} = p_i. \tag{2.52}$$

This result will be useful below.

[5] We shall see that in Quantum Mechanics, this is no longer true.

2.4 Legendre Transformation

Let $f = f(x, y)$ represent the law of a physical phenomenon, with f a good function of two variables that represent measurable quantities. Then, its differential is

$$df = udx + vdy, \tag{2.53}$$

where

$$u = \frac{\partial f}{\partial x}, v = \frac{\partial f}{\partial y}. \tag{2.54}$$

Now suppose there is a one-to-one correspondence between x and u; this is granted if $u = \frac{\partial f(x,y)}{\partial x}$ increases (or decreases) with increasing x. Now we might wish to reformulate the theory in terms of u since u is another measurable quantity. This reformulation may be advantageous (for example, u may be easier to measure than x); indeed, having alternatives is generally preferable, even if the reasons may not be trivial, as we will see in the next paragraphs. Since f depends on x, y we need another function which contains the same information,

It occurred to Legendre[6] that the x dependence is removed if one introduces a new function

$$g = f - ux, \tag{2.55}$$

and treats u and y as the independent variables, because then

$$dg = df - udx - xdu = vdy - xdu. \tag{2.56}$$

The differential dx has disappeared, by the rules of differential calculus. The knowledge of $g(u, y)$ is fully equivalent to the knowledge of $f(x, y)$, and it is a matter of convenience as to which function to use. The inverse transformation is just $f = g + xu$. As in the case of Fourier, if you apply the transform twice, you are back at the original form.

Legendre's method is simple and general and can be extended directly to functions of many variables $f(x, y, z, w, \ldots)$ since the transformation can be applied independently to each variable.

[6] Adrien Marie Legendre (Paris 1752–Paris 1833) academic of Sciences and one of the great mathematicians of his age.

2.5 The Hamiltonian

Building on the Lagrangian formulation, William Rowan Hamilton[7] found a new, more effective framework. The Hamiltonian formulation allows us to solve difficult problems and deepens the general understanding of the workings of the theory. The main idea is a change of the independent variables. No longer $q, \dot{q}$, but q_i and $p_i = \frac{\partial L}{\partial \dot{q}_i}$. At first sight, this does not appear to be a real novelty, since for a point mass in Cartesian coordinates velocity and momentum are proportional, $\overrightarrow{p} = m\frac{d\overrightarrow{x}}{dt}$. However, when using other Lagrangian coordinates, there is no such trivial relation between momentum and velocity, and further complications arise when introducing a vector potential (see Sect. 2.5.4). The change from the independent variables $\dot{q}_i$ to p_i is a Legendre transformation

$$H(p, q, t) = \sum_i p_i \dot{q}_i - L(q, \dot{q}, t), \tag{2.57}$$

where $H(p, q, t)$ is the Hamiltonian, which does not depend on the velocities:

$$\frac{\partial H}{\partial \dot{q}_k} = p_k - \frac{\partial L}{\partial \dot{q}_k} = 0. \tag{2.58}$$

If L does not depend on time explicitly, we know from (2.30) that $H = E$ is the energy and is conserved along the physical evolution of the system; then, we can say that *the Hamiltonian is the energy as a function of q and p*. The equations of motion were obtained by Hamilton from the variational principle $\delta S = 0$, with S now a functional of the path written as $p(t), q(t)$:

$$S = \int_{t_1}^{t_2} dt\, L(q, \dot{q}, t) = \int_{t_1}^{t_2} dt \left[\sum_i p_i \dot{q}_i - H(p, q, t) \right]. \tag{2.59}$$

Here, $\dot{q}(t)$ is thought of as a function of $q(t)$ and $p(t)$ along the virtual path. Note that $q(t)$ e $p(t)$ can now be varied independently, while in the Lagrangian formalism, $\dot{q}$ is fixed by $q(t)$). We consider variations

$$q_i(t, \alpha_i) = q_i(t, 0) + \alpha_i \eta_i(t),$$
$$p_i(t, \alpha_i) = p_i(t, 0) + \alpha_i \xi_i(t),$$

where $\eta_i(t), \xi_i(t)$ are arbitrary independent functions. As in the Lagrangian formalism, q_i is fixed at t_1 and t_2,

[7] This early Irish genius (Dublin 1805–Dublin 1865), at the age of 22, went to Trinity College to explain his reformulation of Mechanics, and immediately got a chair. He continued for the rest of his life to produce high-level results; he also worked on the quaternions, which are essentially the spin matrices that we will discuss later on.

$$\eta_i(t_1) = \eta_i(t_2) = 0. \tag{2.60}$$

Note, however, that no such constraint limits ξ_i. In Lagrangian formalism, there was no prescription on $\dot{q}(t_1)$ and $\dot{q}(t_2)$; in the same way, $p(t_1)$ and $p(t_2)$ are left arbitrary. So, (2.59) becomes

$$S = \int_{t_1}^{t_2} dt \left[\sum_i [p_i + \alpha_i \xi_i] [\dot{q}_i + \alpha_i \dot{\eta}_i] - H(p_i + \alpha_i \xi_i, q_i + \alpha_i \eta_i, t) \right].$$

Neglecting terms in α_i^2,

$$\frac{dS}{d\alpha_i} = \int_{t_1}^{t_2} dt \left(p_i \dot{\eta}_i + \dot{q}_i \xi_i - \frac{\partial H}{\partial q_i} \eta_i - \frac{\partial H}{\partial p_i} \xi_i \right) = \int_{t_1}^{t_2} dt \left(p_i \dot{\eta}_i - \frac{\partial H}{\partial q_i} \eta_i + \xi_i \left[\dot{q}_i - \frac{\partial H}{\partial p_i} \right] \right).$$

An integration by parts eliminates $\dot{\eta}$ and we are left with

$$\int_{t_1}^{t_2} dt \left[\eta_i \left(-\dot{p}_i - \frac{\partial H}{\partial q_i} \right) + \xi_i \left(\dot{q}_i - \frac{\partial H}{\partial p_i} \right) \right] = 0.$$

Hence, one obtains the Hamiltonian equations (or *canonical equations*)

$$\dot{p}_i = -\frac{\partial H}{\partial q_i}, \tag{2.61}$$

$$\dot{q}_i = \frac{\partial H}{\partial p_i}. \tag{2.62}$$

The p, q that appear in (2.61), (2.62) are canonically conjugate variables. The canonical equations also follow from Lagrange's equations, but to see that, starting from the definition (2.57) of the Hamiltonian, one must be careful about the independent variables.

For L they are q, $\dot{q}$, but for $H(p, q, t) = \sum_i p_i \dot{q}_i - L(q, \dot{q}, t)$, $\dot{q}$ must be rewritten in terms of p and q. Thus,

$$\frac{\partial H}{\partial q_k} = \sum_i p_i \frac{\partial \dot{q}_i}{\partial q_k} - \frac{\partial L}{\partial q_k} - \sum_i \frac{\partial L}{\partial \dot{q}_i} \frac{\partial \dot{q}_i}{\partial q_k}. \tag{2.63}$$

The first term yields 0 with the third and one gets, using Lagrange's equations,

$$\frac{\partial H}{\partial q_k} - \frac{\partial L}{\partial q_k} = -\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_k} = -\dot{p}_k. \tag{2.64}$$

This verifies (2.61). On the other hand, differentiating (2.57),

$$\frac{\partial H}{\partial p_k} = \dot{q}_k + \sum_i p_i \frac{\partial \dot{q}_i}{\partial p_k} - \sum_i \frac{\partial L}{\partial \dot{q}_i}\frac{\partial \dot{q}_i}{\partial p_k};$$

simplifying the last two terms one is left with (2.62). Along a physical trajectory, $H = H(p(t), q(t), t)$ depends on the time in such a way that

$$\frac{dH}{dt} = \frac{\partial H}{\partial t}. \tag{2.65}$$

Indeed,

$$\frac{dH}{dt} = \frac{\partial H}{\partial t} + \sum_i \left(\frac{\partial H}{\partial q_i}\dot{q}_i + \frac{\partial H}{\partial p_i}\dot{p}_i\right),$$

but the sum vanishes by the canonical equations. So, when H is not explicitly a function of time, then $H = E$ is a constant.

2.5.1 Reduced Mass

For $N = 2$ point masses m_1, m_2 interacting via an instantaneous potential energy $V(r_{12})$ depending on their distance, one can start from the observation that $m_1 \ddot{\vec{r}}_1 = -\nabla_1 V(r_{12})$ and $m_2 \ddot{\vec{r}}_2 = -\nabla_2 V(r_{12})$ are opposite; in other words, the instantaneous forces felt by the two masses are opposite. Consequently, their center of mass

$$\vec{R} = \frac{m_1 \vec{r}_1 + m_2 \vec{r}_2}{m_1 + m_2} \tag{2.66}$$

obeys $\ddot{\vec{R}} = 0$, that is, makes a free-particle motion.

Thus, the Hamiltonian

$$H = \frac{p_1^2}{2m_1} + \frac{p_2^2}{2m_2} + V(r_{12}) \tag{2.67}$$

can be simplified, by separating the center-of-mass degrees of freedom from the more interesting relative motion.

One can solve the problem in terms of a single effective particle having the reduced mass μ such that

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2}. \tag{2.68}$$

Setting $\rho = r_{12}$, $\boldsymbol{p_R} = \boldsymbol{p}_1 + \boldsymbol{p}_2$ and $\boldsymbol{p}_\rho = \mu(\boldsymbol{v}_1 - \boldsymbol{v}_2)$, one can change variables with

$$\overrightarrow{p}_1 = \overrightarrow{p}_{\vec{\rho}} + \frac{m_1}{m_1 + m_2}\overrightarrow{p}_{\vec{R}}, \qquad \overrightarrow{p}_2 = -\overrightarrow{p}_{\vec{\rho}} + \frac{m_2}{m_1 + m_2}\overrightarrow{p}_{\vec{R}}.$$

Computing the squares one realizes that

$$\frac{p_1^2}{m_1} + \frac{p_2^2}{m_2} = \frac{p_R^2}{m_1 + m_2} + \left[\frac{1}{m_1} + \frac{1}{m_2}\right] p^2_{\vec{\rho}}. \tag{2.69}$$

Since $\overrightarrow{p}_{\vec{R}}$ is conserved, we can separate the center-of-mass motion and reduce the two-body problem to an effective one-body problem with Hamiltonian

$$H = \frac{\boldsymbol{p}_\rho^2}{2\mu} + V(\rho).$$

We already used this separation in Sect. 2.2.2. Once the relative motion is determined, one can use $\rho(t)$ and $\boldsymbol{R}(t)$ to find $\boldsymbol{r}_1(t)$ and $\boldsymbol{r}_2(t)$. This argument is easily extended to reduce any N body problem to $N - 1$: for $N = 3$ one can separate the motion of the center of mass

$$\overrightarrow{R} = \frac{m_1\overrightarrow{r}_1 + m_2\overrightarrow{r}_2 + m_3\overrightarrow{r}_3}{m_1 + m_2 + m_3} \tag{2.70}$$

and reduce the problem to 2 effective particles.

2.5.2 *Canonical Transformations*

The Lagrangian formalism allows us to use any choice of the s coordinates q that specify the configuration of the system, and this freedom helps in the setup and solution of equations. The change from one set of coordinates q to a new set Q is a *point transformation*. The change in the velocities implied by the change of the coordinates does not add anything. The Hamiltonian formalism replaces the s second-order Lagrange equations with $2s$ first-order Eqs. 2.61, 2.62 and enlarges the class of transformations from the point transformations to the *canonical transformations*. Such transformations take from coordinates q_i and canonically conjugate momenta $p_i, i = 1 \ldots s$, to new coordinates and momenta Q_i and P_i. In short we may say

that the transformation is $q, p \to Q, P$. Both choices correspond to different pictures or representations of the same physical reality. The new independent variables Q, P are assumed to be in one-to-one correspondence with q, p via some invertible functional dependence such that $P = P(p, q, t)$ and $Q = Q(p, q, t)$, and must be canonically conjugate to a new Hamiltonian $\tilde{H}(P, Q, t)$. Each virtual path $p(t), q(t)$ is also represented by $P(t), Q(t)$, and the physical paths correspond to solutions of the Hamilton equations of motions in both pictures. It is clear that the point transformations are a special case, but Hamilton's formalism is much more rewarding both in the solution of mechanical problems and in the clarification of the inner mathematical structure of the theory.

It is desirable to develop a general technique that, starting from $H(p, q, t)$, generates canonical transformations to $P(p, q, t)$, $Q(p, q, t)$ and the corresponding new Hamiltonians $\tilde{H}(P, Q, t)$. If we assume that the Lagrangian is unaffected by the transformation, that is,

$$\sum_i p_i \dot{q}_i - H(p, q, t) \equiv \sum_i P_i \dot{Q}_i - \tilde{H}(P, Q, t),$$

we are granted that the action is the same for each virtual path and the condition $\delta S = 0$ gives the same paths in the two representations. This is correct, but too restrictive, since the Lagrangian is largely arbitrary, and we can allow it to be different in the two pictures. What we really want is that a path which satisfies $\delta S = 0$ in the (p, q) representation is mapped to a minimal path in the (P, Q) representation. In other words, we just need that for all the variations with fixed end points of any virtual path,

$$\delta S = \delta \int_{t_1}^{t_2} dt \left[\sum_i p_i \dot{q}_i - H \right] = \delta \int_{t_1}^{t_2} dt \left[\sum_i P_i \dot{Q}_i - \tilde{H} \right]. \tag{2.71}$$

The condition (2.71) is fulfilled if the integrands differ by the total derivative $\frac{dF}{dt}$ of a function $F = F(p, q, P, Q, t)$. The total derivative inserted into the integral gives some constant, and then δ yields 0. F is called the *generating function of the canonical transformation*.

Indeed,

$$\delta \int_{t_1}^{t_2} dt \frac{dF}{dt} = \delta[F(t_2) - F(t_1)] = 0.$$

Evidently, $F = F(p, q, P, Q, t)$ is an action.

Therefore, a condition that ensures a canonical transformation is

$$\sum_i p_i \dot{q}_i - H = \sum_i P_i \dot{Q}_i - \tilde{H} + \frac{dF}{dt}. \tag{2.72}$$

Here, p, q are functions of t along a path, while $Q(t)$, $P(t)$ are a different representation of the same path. A clever choice of F can allow for drastic simplifications of hard problems. However we cannot be too euphoric: there is no method for finding F, and it is difficult to invent a function of many variables. So in a sense the method is too general to be practical. Normally one assumes that $F = F(p, q, P, Q, t)$ does not really depend on all its variables. For instance, $F = F(q, Q, t)$ leads to

$$\sum_i p_i \dot{q}_i - H = \sum_i P_i \dot{Q}_i - \tilde{H} + \frac{\partial F}{\partial t} + \sum_i [\frac{\partial F}{\partial q_i}\dot{q}_i + \frac{\partial F}{\partial Q_i}\dot{Q}_i]. \tag{2.73}$$

The coefficients of $\dot{q}_i$ on both sides must be equal,

$$p_i = \frac{\partial F(q, Q, t)}{\partial q_i}. \tag{2.74}$$

The same argument applies to $\dot{Q}_i$:

$$P_i = -\frac{\partial F(q, Q, t)}{\partial Q_i}. \tag{2.75}$$

The transformation is canonical provided that

$$\tilde{H} = H + \frac{\partial F(q, Q, t)}{\partial t}. \tag{2.76}$$

Thus, unlike the Lagrangian, $\tilde{H}$ is not obtained from H with a simple change of variables; $\tilde{H} = H$ only if F is time-independent. Any $F(q, Q, t)$ that one can invent will lead to some $\tilde{H}$ (although a smart choice is needed to find a useful transformation); however, we are not finished, since $\tilde{H}$ deserves to be considered the new Hamiltonian only after eliminating p, q in favor of P, Q. Thus, we must be able to do the transformation $(p, q) \to (P, Q)$. We must be able to solve Eqs. 2.74 and 2.75 for P e Q in terms of p, q. Once we know $P(p, q, t)$ and $Q(p, q, t)$ and the inverse functions $p(P, Q, t)$ and $q(P, Q, t)$, then $\tilde{H}$ is achieved and we have a transformed problem to solve. If we can solve $\tilde{H}$, then the transformation $(P, Q) \to (p, q)$ solves the original problem.

Example 2.1 The generating function

$$F = \sum_i \lambda_i q_i Q_i \tag{2.77}$$

yields

$$p_i = \lambda_i Q_i, \quad P_i = -\lambda_i q_i. \tag{2.78}$$

There is no reason to talk about momenta and coordinates any longer. One would do better to speak about canonically conjugate variables. (Dimensionally pq is an action, and this remains true, while the dimensions of the coordinates and momenta may change.) For example, a harmonic oscillator is sent by the transformation into a harmonic oscillator.

Problem 2.8 Let

$$H(p,q,t) = \alpha pt + \mu\left(q - \frac{1}{2}\alpha t^2\right)^2, \tag{2.79}$$

with constant α, μ. Solve the mechanical problem with the canonical transformation specified by the generating function

$$F(q, Q, t) = Q\left(q - \frac{1}{2}\alpha t^2\right). \tag{2.80}$$

Solution 2.8 One finds

$$p = Q,\ P = -\left(q - \frac{1}{2}\alpha t^2\right),\ \frac{\partial F}{\partial t} = -\alpha t Q,$$

and so

$$\tilde{H}(P, Q, t) = \mu P^2.$$

Every motion described by (2.79) is the image of a free motion.

Example 2.2 An important kind of transformation is generated by

$$F = S(q, P, t) - \sum_i P_i Q_i. \tag{2.81}$$

Now it is S which is commonly called the generating function. Then, (2.72) becomes

$$\begin{aligned}\sum_i p_i \dot{q}_i - H = &\sum_i P_i \dot{Q}_i - \tilde{H} + \frac{\partial S}{\partial t} \\ &+ \sum_i \left[\frac{\partial S}{\partial q_i}\dot{q}_i + \frac{\partial S}{\partial P_i}\dot{P}_i - P_i\dot{Q}_i - Q_i\dot{P}_i\right].\end{aligned} \tag{2.82}$$

Equating the coefficients of $\dot{q}_i$ on both sides, one finds

$$p_i = \frac{\partial S}{\partial q_i}. \tag{2.83}$$

this is just Eq. (2.52).

S is the action as a function of the old coordinates and the new momenta at the upper integration limit. Equating the coefficients of one gets

$$Q_i = \frac{\partial S}{\partial P_i},$$
$$\tilde{H} = H + \frac{\partial S}{\partial t}.$$

Example 2.3 $S = \sum_k q_k P_k$, $F = S(q, P, t) - \sum_i P_i Q_i$ generates the identity transformation $P = p$, $Q = q$.

Problem 2.9 Given

$$H(p, q) = vp + \left(\frac{A}{p}\right)^2 + B^2 p^4 (q - vt)^2,$$

where A, B, v are constants, transform the problem according to the generating function

$$F(q, Q, t) = \frac{q - vt}{Q}.$$

Solution 2.9 Since $p = \frac{\partial F}{\partial q} = \frac{1}{Q}$, $P = -\frac{\partial F}{\partial Q} = \frac{q - vt}{Q^2}$, $\frac{\partial F}{\partial t} = -\frac{v}{Q}$, one finds

$$\tilde{H}(P, Q) = H + \frac{\partial F}{\partial t} = A^2 Q^2 + B^2 P^2,$$

which describes a harmonic oscillator with pulsation $\omega = A$ and mass $m = \frac{1}{2B^2}$.

2.5.3 Hamilton–Jacobi Equation

We have just seen in Example 2 that a canonical transformation with a generating function

$$F = S(q, P, t) - \sum_i P_i Q_i \tag{2.84}$$

leads to a transformed Hamiltonian

$$\tilde{H} = H + \frac{\partial S}{\partial t}.$$

Moreover $p_i = \frac{\partial S}{\partial q_i}$.

Here we have the freedom to choose the most convenient Hamiltonian; Hamilton himself (with Jacobi) made an extreme choice. They chose

$$\tilde{H} = 0. \tag{2.85}$$

This is feasible. Indeed, $S = S(q, P, t)$ (function of the old coordinates and the new momenta) can always be chosen such that

$$\tilde{H}(P, Q, t) = H(p, q, t) + \frac{\partial S}{\partial t} = 0. \tag{2.86}$$

Moreover, this implies that if H does not depend explicitly on the time, the generating function is of the form

$$F(q, P, t) = S(q) - Et, \tag{2.87}$$

with constant E. We may note that $\frac{\partial F}{\partial E}$ is a constant time, which will be useful below. Then, the canonical equations $\frac{\partial P}{\partial t} = -\frac{\partial}{\partial Q}\tilde{H}(P, Q, t) = 0$ imply that P are constant and

$$\frac{\partial Q}{\partial t} = -\frac{\partial}{\partial P}\tilde{H}(P, Q, t) = 0 \tag{2.88}$$

imply that Q are constant, too.[8]

Putting $p_i = \frac{\partial S}{\partial q_i}$ into (2.86), one arrives at the celebrated Hamilton–Jacobi equation:

$$H\left(\frac{\partial S}{\partial q}, q, t\right) + \frac{\partial S}{\partial t} = 0. \tag{2.89}$$

This is a first-order partial differential equation in n+1 variables which are the time and the spatial coordinates q_i. To solve, one must find S as a function of the coordinates and the time, with the momenta as constants of the motion. Actually, the unknown S is an old acquaintance. From the differential

$$dS = \sum_i \frac{\partial S}{\partial q_i} dq_i + \frac{\partial S}{\partial t} dt$$

and from $p_i = \frac{\partial S}{\partial q_i}$, one obtains

$$\frac{dS}{dt} = \sum_i p_i \frac{dq_i}{dt} + \frac{\partial S}{\partial t}.$$

Substituting $\frac{\partial S}{\partial t}$ taken from the Hamilton–Jacobi equation, one concludes that

$$\frac{dS}{dt} = \sum_i p_i \frac{dq_i}{dt} - H = L,$$

[8] Please, don't be fooled by the compact notation: recall that, like above, Q stands for the set of coordinates and P for set of the impulses.

that is, S is just the action. If H does not depend on time, (2.86) ensures that $\frac{\partial S}{\partial t} = -E$ is a constant, and then

$$H(\frac{\partial S}{\partial q}, q) = E. \tag{2.90}$$

Here we know how H depends on the momenta, while the dependence of S on q and t is unknown.

The mechanical problem is reformulated in terms of a partial differential equation, like the propagation of a nonlinear wave. This is an important independent method of solution and something of a premonition of Quantum Mechanics. However the methods is surprisingly slow when applied to simple problems like a particle is a constant gravity field or the harmonic oscillator.

Problem 2.10 Solve the problem of the harmonic oscillator, with the Hamiltonian (in obvious notation)

$$H(p, q) = \frac{p^2 + A^2 q^2}{2\,m},$$

by the Hamilton–Jacobi method.

Solution 2.10 Since $H(p, q)$ is time-independent, we use the above form (2.89)

$$H(\frac{\partial S}{\partial q}, q) = E$$

of the Hamilton–Jacobi equation, and obtain

$$(\frac{\partial S}{\partial q})^2 + A^2 q^2 = 2\,mE.$$

This gives

$$\frac{\partial S}{\partial q} = \sqrt{2mE - A^2 q^2}$$

and the generating function (2.5.3) is

$$F = S - Et = \int \sqrt{2mE - A^2 q^2} dq - Et.$$

Now, in order to obtain the time dependence, we use equation (2.5.3) and set $\beta = \frac{\partial F}{\partial E}$. This gives

$$\beta = m \int \frac{1}{\sqrt{2mE - A^2 q^2}} dq - t.$$

Let $x = \frac{Aq}{\sqrt{2mE}}$. Now,

$$\beta + t = \frac{m}{A} \arcsin(\frac{Aq}{\sqrt{2mE}}).$$

Finally, working out $q(t)$ from this equation, one obtains the expected result:

$$q(t) = \frac{\sqrt{2mE}}{A} \sin\left[\frac{A(t+\beta)}{m}\right].$$

This example shows that using Hamilton–Jacobi's method to solve trivial problems is like using a hammer to crush peanuts. This is the price to pay for its great power and generality, provided that we are able to separate the variables. This is very important, because otherwise nonlinear partial differential equations are, generally speaking, much harder to solve analytically than ordinary differential equations. The Hamilton–Jacobi method is powerful in cases when the variables can be separated, that is, it is possible to write the action in the form

$$S = \sum_k S_k(q_k, \alpha_1, \cdot, \alpha_s) - E(\alpha_1, \cdot, \alpha_s)t, \tag{2.91}$$

where $\alpha_1, \cdot, \alpha_s$ is a set of constants, while S_k and E are functions. This leads to the solution also in cases where the other methods based on the use of cyclic variables fail.

2.5.4 Point Charge in an Electromagnetic Field

A point charge q of mass m in an electromagnetic field $\vec{E}$, $\vec{B}$ feels the Lorentz force. In the MKSA system, this reads as:

$$\vec{F} = q\left[\vec{E} + \vec{v} \wedge \vec{B}\right]. \tag{2.92}$$

In the Gauss cgs system,[9] one writes

$$\vec{F} = q\left[\vec{E} + \frac{\vec{v}}{c} \wedge \vec{B}\right].$$

One should not be confused with the signs. Suppose you draw a x,y coordinate system on this page, and take the z axis and the magnetic field pointing up, then a

[9] This is popular in scientific literature; see, e.g., John D. Jackson, "Classical Electrodynamics", John Wiley and Sons (1962).

positive charge with a positive velocity along the x axis should be deflected down (i.e., toward negative y), while an electron would be deflected up. In this section, we do the calculations in the International System. Instead of working with 2 fields (6 components), it is better to express everything in terms of potentials $\vec{A}$ and ϕ (4 components overall); the magnetic induction field is

$$\vec{B} = \vec{\nabla} \wedge \vec{A} = (\partial_2 A_3 - \partial_3 A_2, \partial_3 A_1 - \partial_1 A_3, \partial_1 A_2 - \partial_2 A_1),$$

and the electric field comes from

$$\vec{E} = -\vec{\nabla}\phi - \frac{\partial \vec{A}}{\partial t}; \tag{2.93}$$

so, the equation of motion becomes

$$m\ddot{\vec{r}} = -q\vec{\nabla}\phi - q\frac{\partial \vec{A}}{\partial t} + q\dot{\vec{r}} \wedge \vec{B}. \tag{2.94}$$

We deduced the Euler–Lagrange equations under the hypothesis that force was the gradient of a potential. The Lorentz force depends on the velocity, but don't worry: we will immediately see that a Lagrangian also exists in this case; Eq. (2.94) is obtained from the Lagrangian[10]

$$L(\vec{r}, \vec{v}, t) = \frac{1}{2}mv^2 - q\phi + q\vec{v} \cdot \vec{A}(\vec{r}, t). \tag{2.95}$$

Using Cartesian coordinates, one finds

$$\frac{\partial L}{\partial x_i} = -q\frac{\partial \phi}{\partial x_i} + q\sum_j \dot{x}_j \frac{\partial A_j}{\partial x_i}, \tag{2.96}$$

while

$$p_i = \frac{\partial L}{\partial \dot{x}_i} = m\dot{x}_i + qA_i. \tag{2.97}$$

Since $\vec{A}(\vec{r}, t)$ must be understood as $\vec{A}(\vec{r}(t), t)$, where $\vec{r}(t)$ travels along the trajectory,

$$\dot{\vec{A}} = \left(\frac{\partial}{\partial t} + \dot{x}_1\frac{\partial}{\partial x_1} + \dot{x}_2\frac{\partial}{\partial x_2} + \dot{x}_3\frac{\partial}{\partial x_3}\right)\vec{A}.$$

[10] Any other Lagrangian that differs by a total derivative $\frac{dF(\vec{r},t)}{dt}$ is equivalent.

Therefore the Lagrange equations of motion are

$$m\ddot{x}_i = q\left[-\frac{\partial\phi}{\partial x_i} - \frac{\partial A_i}{\partial t} + \sum_j \dot{x}_j\left(\frac{\partial A_j}{\partial x_i} - \frac{\partial A_i}{\partial x_j}\right)\right]. \tag{2.98}$$

One can readily check that, expanding (2.94) in components, the result is the same. Since the force depends on the velocity, now the Lagrangian is $L = T - U$, with $T = \frac{1}{2}mv^2$ and $U = U(q, \dot{q}, t)$. There is a velocity-dependent generalized force

$$Q_i = -\frac{\partial U}{\partial q_i} + \frac{d}{dt}\frac{\partial U}{\partial \dot{q}_i}, \tag{2.99}$$

which is just the Lorentz one, and x_i is canonically conjugate to the *canonical momentum*

$$p_i = \frac{\partial L}{\partial \dot{x}_i} = m\dot{x}_i + qA_i,$$

which, however, is not observable (like A_i). It is the *mechanical momentum* $p^{mec} \equiv m\dot{x}_i$ that is, of course, observable. The energy

$$\begin{aligned} E &= \textstyle\sum_i p_i\dot{q}_i - L \\ &= \textstyle\sum_i \left[(m\dot{x}_i + qA_i)\dot{x}_i - \left(\frac{1}{2}m\dot{x}_i^2 + q\dot{x}_i \cdot A_i\right)\right] + q\phi = \frac{1}{2}m\dot{r}^2 + q\phi \end{aligned} \tag{2.100}$$

does not depend on the vector potential; however, since it contains ϕ, one can measure neither, but one can measure the difference of ϕ between two space–time points and energy differences. However, the Hamiltonian H must be a function of the canonical momentum $\overrightarrow{p}$. Therefore, the Hamiltonian is

$$H = \frac{1}{2m}(\overrightarrow{p} - q\overrightarrow{A})^2 + q\phi. \tag{2.101}$$

Thus, H contains $\boldsymbol{p}$ and $\boldsymbol{A}$, which are not measurable, while the energy is defined up to an arbitrary constant. To sum up, the rule to include the electromagnetic field in a problem is

$$\boxed{E \to E + q\phi\,,\ \overrightarrow{p} \to \overrightarrow{p} - q\overrightarrow{A}\,.}$$

This rule (*minimal coupling*) remains true in relativistic, quantum and quantum relativistic theories.

Remark 2.1 In the Gauss system, the rule becomes

$$E \to E + q\phi,\ \ \overrightarrow{p} \to \overrightarrow{p} - \frac{q}{c}\overrightarrow{A}\,.$$

2.5.5 Poisson Brackets

Let $F = F(p(t), q(t), t)$ be a measurable physical quantity, a function of the canonical variables; the values it takes at time t while $q(t)$ and $p(t)$ vary along a trajectory must, of course, be invariant under all canonical transformations. The total time derivative of F is

$$\dot{F} = \frac{\partial F}{\partial t} + \sum_i \left(\frac{\partial F}{\partial q_i} \dot{q}_i + \frac{\partial F}{\partial p_i} \dot{p}_i \right).$$

Along the trajectory, $\dot{q} = \frac{\partial H}{\partial p}$, $\dot{p} = -\frac{\partial H}{\partial q}$, and one finds that

$$\dot{F} = \frac{\partial F}{\partial t} + \sum_i \left(\frac{\partial F}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial F}{\partial p_i} \frac{\partial H}{\partial q_i} \right) = \frac{\partial F}{\partial t} + \{F, H\}_{p,q} . \tag{2.102}$$

In the last term, I have introduced the Poisson bracket; for any two functions $A(p, q)$ and $B(p, q)$ of the canonical variables, they are defined by

$$\{A, B\}_{p,q} \equiv \sum_k \left(\frac{\partial A}{\partial q_k} \frac{\partial B}{\partial p_k} - \frac{\partial A}{\partial p_k} \frac{\partial B}{\partial q_k} \right) = - \{B, A\}_{p,q} . \tag{2.103}$$

The symbol $\{A, B\}_{p,q}$ emphasizes that the derivatives are taken with respect to the set (p, q). If F does not depend on time explicitly and $\{F, H\}_{p,q} = 0$, then F is a constant of the motion. Anyhow, since $\dot{F} - \frac{\partial F}{\partial t}$ is a measurable, physical quantity, it must be invariant, i.e., $\{F, H\}_{p,q} = \{F, H\}_{P,Q}$. The Poisson brackets with the Hamiltonian are invariant.

If $H(p, q)$, which governs the time evolution, is replaced by any arbitrary function of the same variables, the invariance of the Poisson bracket under canonical transformation remains. The transformation $\{A, B\}_{p,q} \to \{A, B\}_{P,Q}$ involves writing (p, q) in terms of (P, Q) and then differentiating with respect to the new variables. In this respect, H is not special. One could always consider B as a fictitious Hamiltonian and justify the invariance as above. Therefore, we may change notation and simply write $\{A, B\}$.

The following rules are readily seen:

$$\{A, \text{costant}\} = 0; \tag{2.104}$$

$$\{A, B + C\} = \{A, B\} + \{A, C\}; \tag{2.105}$$

$$\{A, BC\} = \{A, B\} C + \{A, C\} B. \tag{2.106}$$

Moreover, since p and q are independent variables, and

$$\frac{\partial p_i}{\partial p_k} = \delta_{i,k}, \quad \frac{\partial q_i}{\partial q_k} = \delta_{i,k}, \quad \frac{\partial p_i}{\partial q_k} = 0,$$

one finds the *fundamental brackets*

$$\{q_i, p_j\} = \delta_{i,j}, \quad \{q_i, q_j\} = 0, \quad \{p_i, p_j\} = 0. \tag{2.107}$$

Now consider a canonical transformation to new variables $(p, q) \to (P, Q)$. The Hamiltonian $H(p, q)$ is transformed to a new Hamiltonian $\tilde{H}$. Then, P, Q have the fundamental brackets $\{Q_i, P_j\}_{P,Q} = \delta_{i,j}$, where $\{\}_{P,Q}$ implies differentiation with respect to the new variables. The fact that Poisson's brackets are invariant under canonical transformations implies that $\{Q_i, P_j\}_{p,q} = \delta_{i,j}$, and this is of primary practical importance; one can check the fundamental brackets using the old variables, and certify the canonical transformation without knowing the generating function, which remains unknown in most cases. If the problem is time-independent, one obtains $\tilde{H}$ as well by a direct change of variables.

The Poisson brackets also help to find new constants of the motion using a theorem credited to Poisson. One can show[11] that if f and g are integrals of the motion (i.e., constant quantities), then $\{f, g\}$ is also an integral of the motion. It could turn out to be a constant or some combination of f and g, but sometimes it is a useful *new* conserved quantity. The proof is based on the Jacobi[12] identity

$$\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{g, f\}\} = 0. \tag{2.108}$$

Example 2.4 Consider the harmonic oscillator with Hamiltonian

$$H(p, q) = \frac{p^2}{2m} + \frac{1}{2} m\omega^2 q^2; \tag{2.109}$$

the equations of motion are solved by

$$q(t) = \sqrt{\frac{2A}{m\omega}} \sin(\omega t), \quad p = m\dot{q} = \sqrt{2Am\omega} \cos(\omega t).$$

Here, A is the amplitude of oscillation. The energy is $E = H(p, q) = A\omega$.

Through a canonical transformation, one obtains a new Hamiltonian

$$\tilde{H}(A, \phi) = A\omega, \tag{2.110}$$

[11] L. D. Landau e E. M. Lifšits, Mechanics (Pergamon Press) Chap. 7.

[12] Carl Jacobi (Postdam 1804–Berlin 1851) was professor of Mathematics at Königsberg.

where $\phi = \omega t$; so, the new variables are ϕ, A, with

$$A = \frac{p^2}{2m\omega} + \frac{1}{2}m\omega q^2.$$

Is this transformation canonical? One can verify this property by the Poisson brackets, using the fact that

$$\tan(\phi) = \frac{mq\omega}{p}. \tag{2.111}$$

Indeed,

$$\frac{\partial A}{\partial q} = m\omega q, \qquad \frac{\partial A}{\partial p} = \frac{p}{m\omega},$$
$$\frac{\partial \phi}{\partial q} = \frac{m\omega/p}{1+(m\omega q/p)^2}, \quad \frac{\partial \phi}{\partial p} = \frac{-m\omega q/p^2}{1+(m\omega q/p)^2}$$

and

$$\{\phi, A\} = \frac{\partial \phi}{\partial q}\frac{\partial A}{\partial p} - \frac{\partial \phi}{\partial p}\frac{\partial A}{\partial q} = 1. \tag{2.112}$$

The coordinate is ϕ. The canonical equations of motion read as $\dot{A} = 0,\ \dot{\phi} = \omega$.

2.6 Delaunay Elements

We have just seen that the oscillator Hamiltonian (2.109) can be canonically transformed to the amazingly simple form $A\omega$ where A is an action, and $\omega = \dot{\phi}$, where ϕ is an angle variable. This property can be extended to any periodic motion in 1 dimension. Indeed, the equation $H(p, q) = E$ can be solved to find the contour C in phase space defined by $p = p(q, E)$. The action variable is then

$$A = \oint_C pdq, \tag{2.113}$$

where $pdq > 0$ along C, while the contour goes from a minimum $q = qmin$ to a maximum $q = qmax$ and back. Within C, $E > H$, so

$$A = \iint dpdq\theta(E - H(p, q)). \tag{2.114}$$

Thus,

$$\frac{\partial A}{\partial E} = \iint dpdq\delta(E - H(p,q) = \int_{qmin}^{qmax} \left\{ \left[\frac{dq}{|\frac{\partial H}{\partial p}|} \right]_{p>0} + \left[\frac{dq}{|\frac{\partial H}{\partial p}|} \right]_{p<0} \right\}. \tag{2.115}$$

The Hamilton equation says that $\frac{\partial H}{\partial p} = \dot{q}$, and from each dq, one gets two positive contributions $\frac{dq}{|\dot{q}|}$. One can write

$$\frac{dA}{dE} = \oint \frac{dq}{\dot{q}} = \oint dt = T, \tag{2.116}$$

where T is the period. One can conclude that the frequency is

$$\nu = \frac{\partial E}{\partial A}, \tag{2.117}$$

which is a new Hamilton equation, with $E = E(A)$ as the new Hamiltonian and A the transformed momentum. Let ϕ denote the new coordinate, canonically conjugated to A; Hamilton's equations read as:

$$\frac{\partial E}{\partial A} = \dot{\phi} = \nu \tag{2.118}$$

$$\frac{\partial E}{\partial \phi} = 0. \tag{2.119}$$

This technique also lends itself to the solution of periodic motions with more than one degree of freedom, and, in this case, is easy to use provided that one can separate the variables. Then, for each (p, q) pair, there is a separate contribution $A_q = \oint pdq$ to the action. For the Kepler problem, which is separable, the Lagrangian is

$$L = \frac{1}{2}m\left[\dot{r}^2 + r^2\dot{\phi}^2\right] + G\frac{Mm}{r}, \tag{2.120}$$

and since ϕ is cyclic, $p_\phi = mr^2\dot{\phi}$ is just the angular momentum and is a constant of the motion. Hence, $A_\phi = \oint p_\phi d\phi = 2\pi p_\phi$. The Hamiltonian is

$$H = \frac{p_r^2}{2\,m} + \frac{p_\phi^2}{2mr^2} - G\frac{Mm}{r} \tag{2.121}$$

where, for an orbit with $H = E$,

$$p_r = m\dot{r} = \pm\sqrt{2mE + 2G\frac{Mm}{r} - \frac{p_\phi^2}{r^2}}. \tag{2.122}$$

So,

$$A_r = \oint p_r dr \tag{2.123}$$

is the area inside the curve in the (r, p_r) plane of equation $p_r^2 = -\frac{p_\phi^2}{r^2} + 2m\left(E + \frac{GmM}{r}\right)$. Solving this simple problem, it turns out that

$$A_r = -2\pi p_\phi + \pi GMm\sqrt{\frac{2m}{-E}}. \tag{2.124}$$

Solving for E, one finds the new Hamiltonian

$$E = -\frac{2\pi^2 G^2 M^2 m^3}{(A_r + A_\phi)^2}. \tag{2.125}$$

The two frequencies $\nu_r = \frac{\partial E}{\partial A_r}$ and $\nu_\phi = \frac{\partial E}{\partial A_\phi}$ are equal, in agreement with the well-known fact that the orbits are closed. This is related to the special form of the potential, which allows for the conservation of the Runge–Lenz vector

$$\boldsymbol{R} = \frac{\boldsymbol{p} \wedge \boldsymbol{L}}{m} - GmM\frac{\boldsymbol{r}}{r} \tag{2.126}$$

which is pinned at the aphelion–perihelion direction (except for circular orbits, when $\boldsymbol{R}$ vanishes). In order to verify that, one must compute the time derivative of the two terms of $\boldsymbol{R}$ using $\frac{d}{dt}\frac{\boldsymbol{r}}{r} = \frac{r^2\boldsymbol{v} - \boldsymbol{r}(\boldsymbol{r}.\boldsymbol{v})}{r^3}$, where $\boldsymbol{v} = \frac{d\boldsymbol{r}}{dt}$, and the equation of motion. The French astronomer and mathematician Charles-Eugene Delaunay (1816–1872) demonstrated the usefulness of this method in the study of planetary motion. In addition, it is always possible to solve *integrable systems*. By the term integrable, one refers to systems having N degrees of freedom and N constants of the motion I_k such that all the Poisson brackets vanish, i.e., $\{I_j, I_k\} = 0$. One would say that the conserved quantities are in involution.

2.7 Noether Theorem

We have noted that if $\frac{\partial L}{\partial q} = 0$, then the coordinate q is *cyclic,* and the Lagrange equation says that the corresponding momentum (which is called the conjugate momentum) is conserved. This result admits an important generalization about systems that have continuous symmetries. This means that there is a family of invertible transformations, such as translations or rotations, that leave the equations of motion

unchanged. It is always possible to define a product of transformations as the result of applying the first after the second, and technically the family of symmetry transformation is always a Group, but we shall not use any Group property below. For notational convenience, I shall discuss this for one coordinate q, since the extension to several ones is trivial. Consider the symmetry transformation that changes the virtual paths as follows:

$$T_\alpha : q(t) \to Q[q(t), \alpha];$$

for $\alpha \to 0$ to the varied path $Q[q(t), \alpha]$ is close to $q(t)$. In the action integral,

$$L(q, \dot{q}, t) \to L(Q, \dot{Q}, t) = L(q, \dot{q}, t) + \alpha \left(\frac{\partial L}{\partial q} \frac{\partial Q}{\partial \alpha} + \frac{\partial L}{\partial \dot{q}} \frac{\partial \dot{Q}}{\partial \alpha} \right).$$

Since $\dot{Q} = \frac{\partial Q}{\partial q} \dot{q}$, we must insert

$$\frac{\partial \dot{Q}}{\partial \alpha} = \frac{\partial^2 Q}{\partial q \partial \alpha} \dot{q},$$

and since we are varying a physical path, we can replace $\frac{\partial L}{\partial q}$ by $\frac{d}{dt} \frac{\partial L}{\partial \dot{q}}$.

To sum up, an infinitesimal transformation changes L by

$$\Delta L = \alpha \left[\left(\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} \right) \frac{\partial Q}{\partial \alpha} + \frac{\partial L}{\partial \dot{q}} \frac{d}{dt} \frac{\partial Q}{\partial \alpha} \right] = \alpha \frac{d}{dt} \left(\frac{\partial Q}{\partial \alpha} \frac{\partial L}{\partial \dot{q}} \right). \tag{2.127}$$

So, the Lagrangian changes by a total derivative and the equations of motion are not modified by a generic infinitesimal transformation; however, the path is transformed by T_α, and so the action along the path $q(t)$ changes by

$$\Delta S = \alpha \int_{t_1}^{t_2} dt \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \frac{\partial Q}{\partial \alpha} \right) = \Lambda(t_2) - \Lambda(t_1), \tag{2.128}$$

where

$$\Lambda(t) = p(t) \frac{\partial Q(t)}{\partial \alpha} \tag{2.129}$$

and $p = \frac{\partial L}{\partial \dot{q}}$ is the momentum conjugated with $q = Q(\alpha = 0)$. Since T_α is a symmetry, the transformed path must have identical S as the original path, therefore $\Delta S = 0$ and we may conclude that the $\Lambda(t)$ must be conserved. Any continuous symmetry leads to a conservation law. This reasoning extends to systems with n degrees of freedom, and thus

$$\Lambda(t) = \sum_n p_n(t) \frac{\partial Q_n(t)}{\partial \alpha}. \tag{2.130}$$

It is straightforward to show that a translational symmetry leads to the conservation of momentum. For the motion in a central field, the transformation is a rotation $\phi \to \phi + \alpha, r \to r$ and the rotational symmetry leads to the conservation of angular momentum. The most important application is in field theory, in which the theorem gives the conserved currents.

2.8 Chaos

The great french mathematician Pierre-Simon Laplace was certainly inspired by Classical Mechanics in his influential concept of determinism. He wrote:

We may regard the present state of the universe as the effect of its past and the cause of its future.

Later thinkers have contrasted the use of probability in Quantum Mechanics with the determinism of Classical Mechanics. However, the presence of chaos and uncertainty in the classical events of real life is familiar to everybody as a consequence of the impossibility of exactly knowing the initial conditions. Any dynamical system with n degrees of freedom is described by a set of Hamilton first-order differential equations:

$$\frac{dx_i}{dt} = \phi_i(x_1, x_2, \ldots, x_n). \tag{2.131}$$

The equations of motion for slightly different initial conditions must result from $x_i \to x_i + \Delta x_i$. Using a first-order expansion of the Hamilton equations, we obtain

$$\frac{d\Delta x_i}{dt} = \sum_{j=1}^{n} \frac{\partial \phi_i}{\partial x_j} \Delta x_j. \tag{2.132}$$

Now we wish to know the evolution in time of the deviations Δx_i, and the result depends on the problem. For an integrable system that can be described in terms of independent constants of the motion, all the Δx_i must remain bound. More generally, one might assume that ϕ_i is an analytic function of its arguments and one can approximate $\frac{\partial \phi_i}{\partial x_j} = M_{ij}$ by a constant matrix, at least for small Δx_i. This leads to

$$\Delta x_i = \sum_j (e^{Mt})_{ij} \Delta x_j(0). \tag{2.133}$$

Now, if M has eigenvalues L_i, then e^{Mt} has eigenvalues $e^{L_i t}$; the real parts of L_i are called Lyapunov exponents and lead to an exponential (initial) growth of the deviations from the unperturbed problem. This is the butterfly effect, which makes it impossible to make long-term weather forecasts. However, after a characteristic

time, some sort of instability must appear. For example, a solar system will tend to expel planets after a characteristic time, maybe of the order of billions of years. By contrast, since (as we shall see) in Quantum Theory there are no sharp trajectories, there is no quantum chaos, but there are studies concerning the quantum behavior of classically chaotic systems.

Chapter 3
Dirac's Delta

The Dirac delta is the prototype distribution and is an essential tool of Theoretical Physics.

3.1 Definition of the δ

Two δ's are widely used in theoretical Physics (and in this book, too). One is the Kronecker δ, which takes integer arguments and is defined by

$$\delta_{m,n} = \begin{cases} 1, & m = n, \\ 0, & m \neq n. \end{cases} \tag{3.1}$$

This chapter, however, is devoted to Dirac's δ, which has a continuous variable as an argument. Let us start with the Heavyside[1] θ discontinuous function, also known as the step function, defined by

$$\theta(x) = \begin{cases} 1 \text{ se } x > 0, \\ \frac{1}{2} \text{ if } x = 0, \\ 0 \text{ se } x < 0. \end{cases}$$

[1] Oliver Heaviside (1850–1925) was probably the first to use the δ before Dirac, and the work of George Green also implies the concept. Often the names are not historically fair.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_3

With it we can define a rectangular-shaped peak function, of width 2α,

$$\delta_\alpha(x) = \frac{\theta(\alpha^2 - x^2)}{2\alpha}, \tag{3.2}$$

such that

$$\int_{-\infty}^{\infty} \delta_\alpha(x)dx = 1. \tag{3.3}$$

Now pick an analytic[2] function ϕ and the integral $\int_{-\infty}^{\infty} \delta_\alpha(x)\phi(x)dx$ which gives an average of ϕ for $x \in (-\alpha, \alpha)$ and if α is so small that $\phi(x)$ hardly varies in the interval, the result is close to $\phi(0)$. In other words,

$$\lim_{\alpha\to 0} \int_{-\infty}^{\infty} \delta_\alpha(x)\phi(x)dx = \phi(0).$$

In this way, we have justified the exchange of limit and integration

$$\int_{-\infty}^{\infty} \delta(x)\phi(x)dx = \phi(0), \tag{3.4}$$

where

$$\delta(x) \stackrel{?}{=} \lim_{\alpha\to 0} \delta_\alpha(x). \tag{3.5}$$

As a rule, this exchange of limits is earnestly forbidden; $\lim_{\xi\to 0} \lim_{\lambda\to 0} \frac{\lambda}{\lambda^2+\xi^2} = 0$ while $\lim_{\lambda\to 0} \lim_{\xi\to 0} \frac{\lambda}{\lambda^2+\xi^2}$ blows up. The sign of this fault is that Dirac's delta $\delta(x)$, a very special function, is zero everywhere except that it is ∞ in 0. Actually, it is too singular to meet the minimum of assumptions that are needed to define and handle functions in Analysis. For instance, we should always be able to calculate the square of a function, but $\delta(x)^2$ is meaningless. However, we can define $\delta(x)$ as a distribution.[3] This means that this object exists only to be handled with special rules (that we are going to see) and used to do integrals, and this is the only way it can give numbers at the end of the calculation. Despite such limitations, distributions are so useful in Physics that we cannot simply do without them.

[2] A function f is analytic in 0 if there is an interval including 0 in which the Taylor series converges to f.

[3] This procedure recalls the introduction of the negative numbers by the Chinese and the imaginary numbers by Gerolamo Cardano and Rafael Bombelli in the Italian Renaissance. It is sometimes necessary to invent new objects that break the old rules, and doing that properly can be very rewarding. Thus, in order to be able to differentiate under the integral sign, one introduces the distributions.

There is another way to define Dirac's δ,[4] that is,

$$\delta(x) \equiv \frac{d}{dx}\theta(x), \quad \forall x. \tag{3.6}$$

Again, this does not exist when we work with ordinary functions. However, consider a function $\phi(x)$ that we assumed to be good ($\phi(x) \in C^\infty$) and the ordinary integral

$$\int_{-\infty}^{\infty} dx\phi(x)\theta(x_0 - x) = \int_{-\infty}^{x_0} dx\phi(x).$$

Differentiating with respect to the upper limit,

$$\frac{d}{dx_0}\int_{-\infty}^{\infty} dx\phi(x)\theta(x_0 - x) = \phi(x_0).$$

If we take the liberty to exchange differentiation and integration, we discover that the exchange is correct provided that

$$\int_{-\infty}^{\infty} dx\phi(x)\delta(x - x_0) \equiv \phi(x_0). \tag{3.7}$$

This is an alternative way to arrive at (3.4), starting from (3.6), and implies that

$$\int_{-\infty}^{x} du\delta(u - x_0) = \theta(x - x_0).$$

Evidently, $\delta(x - x_0) = 0$ for $x \neq x_0$, while x_0 *blows up*. Note that $x\delta(x) = 0$. We start to see that it is nice to find a proper way to exchange limits, and this is the reason for introducing the generalized functions, or distributions; the δ is just the prototype. We can think of the δ as the *limit* of sequences of ordinary functions δ_α for $\alpha \to 0$. The meaning is

$$\lim_{\alpha\to 0}\int_{-\infty}^{\infty} dx\phi(x)\delta_\alpha(x - x_0) = \phi(x_0).$$

There are many simple functions, with the property (3.3) that are used as alternative versions of δ_α. One is

[4] P. A. M. Dirac (Bristol 1902–Tallahassee, Florida, USA 1984), English physicist with a Swiss father, wrote the relativistic equation for the electron. Dirac gave many fundamental contributions to Physics, always with a taste for mathematical beauty. Among other achievements, he formulated the theory of the magnetic monopole, predicted the existence of antimatter, proposed the bra–ket notation and invented the technique of Second Quantization. He received the Nobel prize in 1933.

$$\frac{\sin^2\left(\frac{x}{\alpha}\right)}{\pi\left(\frac{x^2}{\alpha}\right)}. \tag{3.8}$$

Here is a list of other $\delta_\alpha(x)$ that are often used:

$$\frac{1}{\pi}\frac{\alpha}{x^2+\alpha^2},\ \frac{-1}{\pi}\mathrm{Im}\frac{1}{x+i\alpha},\ \frac{1}{\alpha\sqrt{\pi}}\exp\left[-\frac{x^2}{\alpha^2}\right],\ \frac{\sin\left(\frac{x}{\alpha}\right)}{\pi x}.$$

From the last example, one can derive an important integral representation of δ. Since

$$\int_{-\frac{1}{\alpha}}^{\frac{1}{\alpha}} e^{iqx}dq = 2\frac{\sin\left(\frac{x}{\alpha}\right)}{x} = 2\pi\delta_\alpha(x),$$

it follows that

$$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{iqx}dq. \tag{3.9}$$

This implies that, given an arbitrary function $f = f(x)$ and its Fourier component

$$F(k) = \int_{-\infty}^{\infty}\frac{dx}{2\pi}e^{ikx}f(x),$$

one gets a one-shot proof of Fourier's theorem

$$\int_{-\infty}^{\infty} dkF(k)e^{-ikx} = \int_{-\infty}^{\infty} dke^{-ikx}\int_{-\infty}^{\infty}\frac{dy}{2\pi}e^{iky}f(y)$$

$$= \int_{-\infty}^{\infty}\frac{dy}{2\pi}f(y)2\pi\delta(x-y) = f(x).$$

The δ is the Fourier transform of $f(x) = 1$. Since $\delta(x)$ vanishes everywhere except at $x = 0$, it is evident that for each interval (a, b),

$$\int_a^b dx\delta(x-x_0) = \begin{cases} 1, & a < x_0 < b, \\ 0, & x_0 \notin (a,b). \end{cases}$$

Since $\delta(x)$ is real, the representation (3.9) implies that

$$\delta(-x) = \delta(x)^* = \delta(x);$$

so, δ is even, and for real a,

$$\delta(ax) = \delta(|a|x);$$

hence, $\int_{-\infty}^{\infty} dx\phi(x)\delta(ax) = \frac{1}{|a|}\int_{-\infty}^{\infty} d(|a|x)\phi(x)\delta(|a|x) = \frac{\phi(0)}{|a|}$, and so,

$$\delta(ax) = \frac{\delta(x)}{|a|}. \tag{3.10}$$

Now let $g(x)$ be a good function with an isolated zero in $x = x_0$; we may write in an interval including x_0 $g(x) \approx \left.\frac{dg}{dx}\right|_{x_0} (x - x_0)$. Assuming that $g'(x_0) \neq 0$,

$$\delta(g(x)) = \delta\left[\left.\frac{dg}{dx}\right|_{x_0} (x - x_0)\right] = \frac{\delta(x - x_0)}{\left|\frac{dg}{dx}\right|}. \tag{3.11}$$

If $g(x)$ has a countable set of zeros at $x = x_\alpha$, $\alpha = 1, 2, \ldots$, we may use the fact that around each zero, $g(x) \approx \left.\frac{dg}{dx}\right|_{x_\alpha} (x - x_\alpha)$. Then,

$$\delta(g(x)) = \frac{\sum_\alpha \delta(x - x_\alpha)}{\left|\frac{dg}{dx}\right|}. \tag{3.12}$$

Some relations that are familiar among ordinary functions must be generalized to the space of distributions. Consider the equation

$$xf(x) = 1. \tag{3.13}$$

If f is a distribution, since $x\delta(x) = 0$, the general solution is

$$f(x) = P\frac{1}{x} + C\delta(x), \quad \text{C} = \text{constant}, \tag{3.14}$$

where P selects the *principal part* of the integral, defined by

$$P\int_{-\infty}^{\infty} \frac{1}{x}\phi(x)dx \equiv \lim_{\epsilon\to 0}\left(\int_{-\infty}^{-\epsilon} + \int_{\epsilon}^{\infty}\right)\frac{1}{x}\phi(x)dx. \tag{3.15}$$

P allows us to tackle some kinds of divergence in 0.

We also need the derivatives of the δ. To find out how

$$\delta'(x) = -\frac{d}{dx}\delta(x) \tag{3.16}$$

works, we integrate by parts

$$\int_{-\infty}^{\infty} dx\delta'(x)\phi(x) = -\int_{-\infty}^{\infty} dx\delta(x)\phi'(x) = -\phi'(0). \tag{3.17}$$

This can be iterated to give the action of the nth derivative,

$$\int_{-\infty}^{\infty} dx \delta^{(n)}(x)\phi(x) = (-1)^n \left.\phi^{(n)}\right|_{x=0}. \tag{3.18}$$

Moreover, the definition of δ has a natural extension to three-dimensional space with

$$\int_{\Omega} d^3 r \delta^{(3)}(\overrightarrow{r} - \overrightarrow{r}_0) = \begin{cases} 1, \overrightarrow{r}_0 \in \Omega, \\ 0, \overrightarrow{r}_0 \notin \Omega. \end{cases} \tag{3.19}$$

In Cartesian coordinates,

$$\delta^{(3)}(\overrightarrow{r} - \overrightarrow{r}_0) \equiv \delta(\overrightarrow{r} - \overrightarrow{r}_0) = \delta(x - x_0)\delta(y - y_0)\delta(x - z_0). \tag{3.20}$$

We can go to curvilinear coordinates by introducing the Jacobian determinant, therefore, while $d^3x \to \left|\frac{\partial(x,y,z)}{\partial(\xi,\eta,\zeta)}\right| d\xi d\eta d\zeta$, the δ must be divided by the same Jacobian:

$$\phi(\overrightarrow{r}_0) = \phi(x_0, y_0, z_0) = \int \phi(\overrightarrow{r})\delta(\overrightarrow{r} - \overrightarrow{r}_0) dx dy dz$$

becomes

$$\begin{aligned} \phi(\overrightarrow{r}_0) &= \phi(\xi_0, \eta_0, \zeta_0) \\ &= \int \phi(\overrightarrow{r}) \left\{ \frac{\delta(\xi - \xi_0)\delta(\eta - \eta_0)\delta(\zeta - \zeta_0)}{\left|\frac{\partial(x,y,z)}{\partial(\xi,\eta,\zeta)}\right|} \right\} \left|\frac{\partial(x,y,z)}{\partial(\xi,\eta,\zeta)}\right| d\xi d\eta d\zeta. \end{aligned} \tag{3.21}$$

For example, in spherical coordinates,

$$\delta(\overrightarrow{r} - \overrightarrow{r}_0) = \frac{\delta(r - r_0)\delta\left[\cos(\theta) - \cos(\theta_0)\right]\delta(\phi - \phi_0)}{r^2}. \tag{3.22}$$

Using the δ, one can define a measure for a *hypersurface* in an N-dimensional space. Take for example for $N = 2$, (xy plane) and consider the family of closed curves defined by

$$f(x, y) = C.$$

The *invariant measure* $\omega(C)$ of a member of the family can be defined by

$$\omega(C) = \int \int dx dy \delta(C - f(x, y)) = \frac{d}{dC}\Omega(C), \tag{3.23}$$

where

$$\Omega(C) = \int\int dxdy\theta(C - f(x, y)) \tag{3.24}$$

is the area inside $f = C$; in general, $\omega(C)$ does not coincide with the length, but is still independent of the system of coordinates.

3.1.1 Volume of the Hypersphere in N Dimensions

The equation of the hypersphere of radius R in the N-dimensional space $\mathbb{R}^N$ is

$$\sqrt{\sum_i^N x_i^2} = R.$$

Its volume

$$\Omega_N(R) = \int_{\sum_i^N x_i^2 < R^2} dx_1 dx_2 \dots dx_N = \int_{\mathbb{R}^N} d^N x\theta(R - r), \tag{3.25}$$

for dimensional reasons, is $\Omega_N(R) = c_N R^N$, where c_N is the volume of the hypersphere with $R = 1$. A neat trick allows us to calculate c_N. Consider the following Gaussian integral:

$$I = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \dots \int_{-\infty}^{\infty} dx_N e^{-\sum_i^N x_i^2} = \left(\int_{-\infty}^{\infty} \mathrm{d}x e^{-x^2}\right)^N = \pi^{\frac{N}{2}}.$$

Putting $\sum_i^N x_i^2 = r^2$, we may write

$$I = \int d^N x e^{-r^2}. \tag{3.26}$$

But $e^{-r^2} = \int_0^\infty dR e^{-R^2}\delta(R - r)$, therefore

$$I = \int d^N x \int_0^\infty dR e^{-R^2}\delta(R - r) = \int_0^\infty dR e^{-R^2} \frac{d}{dR} \int d^N x\theta(R - r).$$

So

$$I = \int_0^\infty dR e^{-R^2} \frac{d}{dR}\Omega_N(R). \tag{3.27}$$

Now the calculation is immediate. Inserting $\Omega_N(R) = c_N R^N$, one obtains

$$\begin{aligned} I &= Nc_N \int_0^\infty dR R^{N-1} e^{-R^2} = Nc_N \int_0^\infty R dR R^{N-2} e^{-R^2} \\ &= \tfrac{Nc_N}{2} \int_0^\infty dt t^{\frac{N-2}{2}} e^{-t} = \tfrac{Nc_N}{2} \left(\tfrac{N}{2} - 1\right)! \end{aligned} \tag{3.28}$$

Here, we met the factorial function

$$x! = \int_0^\infty t^x e^{-t} dt, \tag{3.29}$$

which is $n! = \prod_k^n k$ for integer n, but $\left(\frac{1}{2}\right)! = \frac{\sqrt{\pi}}{2}$, $\left(\frac{3}{2}\right)! = \frac{3\sqrt{\pi}}{4}$ and so on. Equating this to $\pi^{N/2}$, we find

$$\Omega_N(R) = \frac{\pi^{N/2}}{\left(\frac{N}{2}\right)!} R^N. \tag{3.30}$$

Therefore, $\Omega_1 = 2R$, $\Omega_2 = \pi R^2$, $\Omega_3 = \frac{4}{3}\pi R^3$, $\Omega_4 = \frac{1}{2}\pi^2 R^4$, and so on. The measure of the surface of the hypersphere as defined above is $\omega_N(R) = \frac{d}{dR}\Omega_N(R)$.

3.1.2 Plancherel Theorem

Let $\alpha(t), \beta(t)$ be functions belonging to L^2 (such that $\int_{-\infty}^{\infty} |\alpha(t)|^2 dt$ and $\int_{-\infty}^{\infty} |\beta(t)|^2 dt$ exist). Let us denote their Fourier transforms by $\alpha(\omega)$, $\beta(\omega)$. Then,

$$\int_{-\infty}^{\infty} \alpha(t)\beta^*(t) dt = \int_{-\infty}^{\infty} \alpha(\omega)\beta^*(\omega) d\left(\frac{\omega}{2\pi}\right). \tag{3.31}$$

By inserting the definition of the Fourier transforms on the R.H.S., exchanging the order of integration and using the δ functions, one can readily check this important result. One can note that the L.H.S. resembles a scalar product $\sum_i \alpha_i \beta_i^*$, where the discrete index i is replaced by a continuous one, and a similar interpretation is possible for the R.H.S. So, in a sense, this theorem is related to the invariance of scalar products under unitary transformations.

Chapter 4
Some Consequences of Maxwell's Equations

Maxwell completed his field equations in 1864 and published his Treatise in 1873. His equations have not been revised, continuing to be unharmed through the revolutionary changes produced by Relativity and Quantum Mechanics. Nevertheless, the quantum theory of light and the use of lasers have revealed many new phenomena. This might appear to be a contradiction, but it is the plain truth. I prepare the reader to appreciate how this came about by discussing some important facts about classical electromagnetism in this chapter.

4.1 Fields and Potentials

The (classical) electromagnetic fields in vacuo that satisfy given boundary conditions can be calculated through Maxwell's equations. In the Gauss system they read as

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} &= 4\pi\rho, \\ \vec{\nabla} \cdot \vec{B} &= 0, \\ \vec{\nabla} \wedge \vec{E} &= -\frac{1}{c}\frac{\partial \vec{B}}{\partial t}, \\ \vec{\nabla} \wedge \vec{B} &= \frac{1}{c}\frac{\partial \vec{E}}{\partial t} + \frac{4\pi}{c}\vec{j}, \end{aligned} \tag{4.1}$$

where $\vec{j}$ and ρ are current density and charge density. In all, they are 4 functions of space and time. The fields can be computed and measured, however, it amazing that the Maxwell equations succeed in giving us 6 measurable quantities (3 components of $\vec{E}$ and 3 of $\vec{B}$) having only 4 quantities in input. This is a most remarkable property of the electromagnetic field. Moreover, we can obtain the same field more easily by working out 4 quantities, namely the scalar potential ϕ and the vector potential $\vec{A}$, such that

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_4

$$\begin{aligned} \overrightarrow{B} &= \overrightarrow{\nabla} \wedge \overrightarrow{A}, \\ \overrightarrow{E} &= -\overrightarrow{\nabla}\phi - \tfrac{1}{c}\tfrac{\partial \overrightarrow{A}}{\partial t}. \end{aligned} \tag{4.2}$$

From the potentials, one readily obtains the fields, but the potentials are unobservable and largely arbitrary. So, the 4 potentials not only contain all the information about the 6 field components, but also a disposable amount of irrelevant information. This property, called gauge invariance, is a key property of electromagnetism. The choice of the potentials is called *gauge*. In the Lorentz gauge $div\,\overrightarrow{A} + \frac{1}{c}\frac{\partial \phi}{\partial t} = 0$, the Maxwell equations give us

$$\begin{aligned} \left[\nabla^2 - \tfrac{1}{c^2}\tfrac{\partial^2}{\partial t^2}\right]\overrightarrow{A} &= -\tfrac{4\pi}{c}\overrightarrow{J}, \\ \left[\nabla^2 - \tfrac{1}{c^2}\tfrac{\partial^2}{\partial t^2}\right]\phi &= -4\pi\rho. \end{aligned} \tag{4.3}$$

We shall obtain the potentials by solving these equations according to Green's function method. For all $r \neq 0$, $\nabla^2(\frac{1}{r}) = 0$; indeed, for sphericosymmetric functions $\phi(r)$, one can write[1] $\nabla^2\phi(r) = \frac{1}{r}\frac{\partial^2}{\partial r^2}(r\phi(r))$, and since $\phi = \frac{1}{r}$, the result is 0. But this is not the end of the story. Considering a sphere around the origin,

$$\int d^3r\nabla^2\left(\frac{1}{r}\right) = \int d^3r \text{div}\ \ \text{grad}\frac{1}{r} = \int_S \text{grad}\frac{1}{r}\cdot\overrightarrow{n}\,dS$$

and since $\text{grad}(\frac{1}{r}) = -\frac{\overrightarrow{r}}{r^3}$, $\overrightarrow{n} = \frac{\overrightarrow{r}}{r}$, $dS = r^2 d\Omega$, one is left with

$$\int d^3r\nabla^2\left(\frac{1}{r}\right) = -4\pi.$$

In conclusion, $\nabla^2(\frac{1}{r}) = -4\pi\delta(\overrightarrow{r})$. The function $g(r) = \frac{1}{r}$ is called Green's function[2] of the Poisson equation

$$\nabla^2 V(\overrightarrow{x}) = -4\pi\rho(\overrightarrow{x}).$$

[1] An equivalent alternative which is in use is $\nabla^2\phi(r) = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2\frac{\partial}{\partial r}\phi(r))$.

[2] Green's function is named after the Englishman George Green, a solitary amateur genius who working in his mill in the Midlands invented mathematical methods essential for the theory of electromagnetism and all of the modern Theoretical Physics, although at his time Dirac's delta was not yet known.

The solution is the sum of all the potentials of all the point charges that make up $\rho(\overrightarrow{x})$, namely

$$V(\overrightarrow{x}) = \int d^3x' g(\overrightarrow{x} - \overrightarrow{x}')\rho(\overrightarrow{x}').$$

4.2 Green's Function of the Wave Equation and Retarded Potentials

As a prototype calculation illustrating the main steps necessary to compute a Green's function G for many systems of interest, I show in detail how the same reasoning leads to Green's function of the wave equation; by definition,

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right] G(\overrightarrow{r}, t) = -4\pi\delta(\overrightarrow{r})\delta(t). \tag{4.4}$$

Here, an instantaneous localized disturbance in ($\overrightarrow{r} = 0, t = 0$) stands as a source of a potential G that later arrives in ($\overrightarrow{r}, t$). It turns out that G is simply given by $\frac{\delta(t-\frac{r}{c})}{r}$.

To see that, outside the origin, $G(\overrightarrow{r}, t)$ solves the wave equation, we use

$$\nabla^2\phi(r) = \frac{1}{r}\frac{\partial^2}{\partial r^2}(r\phi(r)).$$

For $r \neq 0$,

$$\nabla^2\left(\frac{\delta(t-\frac{r}{c})}{r}\right) = \frac{1}{r}\frac{\partial^2}{\partial r^2}\delta\left(t - \frac{r}{c}\right);$$

now, for each $f(t - \frac{r}{c})$, it holds that

$$\frac{\partial^2}{\partial r^2} f\left(t - \frac{r}{c}\right) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} f\left(t - \frac{r}{c}\right);$$

in particular,

$$\nabla^2\left(\frac{\delta(t-\frac{r}{c})}{r}\right) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\left(\frac{\delta(t-\frac{r}{c})}{r}\right),$$

and

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right] G(\overrightarrow{r}, t) = 0, \quad r \neq 0.$$

To include the effect of the δ, integrate $[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}]G(\overrightarrow{r}, t)$ over a sphere S with any radius R_0 centered on the origin. We denote the surface of the sphere by δS.

Contribution of the term in ∇^2

By the Gauss theorem:

$$\int_S \nabla^2 \left(\frac{\delta(t - \frac{r}{c})}{r} \right) d^3r = \int_{\delta S} \text{grad} \left(\frac{\delta(t - \frac{r}{c})}{r} \right) \cdot \frac{\overrightarrow{r}}{r} dS.$$

In spherical coordinates,

$$\int_S \nabla^2 \left(\frac{\delta\left(t - \frac{r}{c}\right)}{r} \right) d^3r = \int \left[\text{grad} \left(\frac{\delta\left(t - \frac{r}{c}\right)}{r} \right) \cdot \frac{\overrightarrow{r}}{r} \right]_{r=R_0} R_0^2 d\Omega. \tag{4.5}$$

Now, $\frac{\partial}{\partial x} r = \frac{x}{r}$, and so grad $r = \frac{\overrightarrow{r}}{r}$. Hence,

$$\text{grad } \delta\left(t - \frac{r}{c}\right) = -\frac{1}{c} \text{grad } r \; \delta'\left(t - \frac{r}{c}\right) = -\frac{\overrightarrow{r}}{rc} \delta'\left(t - \frac{r}{c}\right),$$

and

$$\text{grad } \frac{\delta\left(t - \frac{r}{c}\right)}{r} = -\frac{\overrightarrow{r}}{r^2 c} \delta'\left(t - \frac{r}{c}\right) = -\frac{\overrightarrow{r}}{r^3} \delta\left(t - \frac{r}{c}\right).$$

Put into (4.5), the angular integral is 4π and

$$\int_S \nabla^2 \left(\frac{\delta\left(t - \frac{r}{c}\right)}{r} \right) d^3r = -4\pi \left\{ \delta\left(t - \frac{R_0}{c}\right) + \frac{R_0}{c} \delta'\left(t - \frac{R_0}{c}\right) \right\}. \tag{4.6}$$

Term in $\dfrac{\partial^2}{\partial t^2}$

Again using $\frac{\partial^2}{\partial r^2}\delta(t - \frac{r}{c}) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\delta(t - \frac{r}{c})$,

$$\begin{aligned} &\textstyle\int_S d^3r \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\left(\frac{\delta(t-\frac{r}{c})}{r}\right) = \int_S d^3r \frac{1}{r}\frac{\partial^2}{\partial r^2}\delta\left(t - \frac{r}{c}\right) \\ &= 4\pi \textstyle\int dr r^2 \frac{1}{r}\frac{\partial^2}{\partial r^2}\delta\left(t - \frac{r}{c}\right). \end{aligned}$$

Integrating by parts, we obtain

$$4\pi \int_0^{R_0} dr \left\{ \frac{\partial}{\partial r} \left[r \frac{\partial}{\partial r} \delta\left(t - \frac{r}{c}\right) \right] - \frac{\partial}{\partial r} \delta\left(t - \frac{r}{c}\right) \right\}.$$

The integrand is a derivative, and one gets

$$= 4\pi \left[r \frac{\partial}{\partial r} \delta \left(t - \frac{r}{c} \right) - \delta \left(t - \frac{r}{c} \right) \right]_{r=0}^{r=R_0}.$$

So,

$$\int_S d^3 r \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \left(\frac{\delta \left(t - \frac{r}{c} \right)}{r} \right) 4\pi \left\{ -\delta \left(t - \frac{R_0}{c} \right) + \frac{R_0}{c} \delta' \left(t - \frac{R_0}{c} \right) + \delta(t) \right\}. \quad (4.7)$$

End result

In conclusion, putting together (4.6) and (4.7),

$$\int_S \left[\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right] G(\overrightarrow{r}, t) d^3 r = -4\pi \delta(t),$$

and so

$$\left[\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right] G(\overrightarrow{r}, t) = -4\pi \delta(\overrightarrow{r}) \delta(t).$$

In summary, $G(\overrightarrow{r}, t) = \frac{\delta(t - \frac{r}{c})}{r}$ is Green's function of the wave equation. Hence, from the geometry of Fig. 4.1 we obtain the potential at $\overrightarrow{x}$ due to a distribution of charges, namely

$$\begin{aligned} \phi(\overrightarrow{x}, t) &= \int d^3 x' \int dt' G(\overrightarrow{x} - \overrightarrow{x}', t - t') \rho(\overrightarrow{x}', t') \\ &= \int d^3 x' \int dt' \frac{\delta(t' - t + \frac{|\overrightarrow{x} - \overrightarrow{x}'|}{c})}{|\overrightarrow{x} - \overrightarrow{x}'|} \rho(\overrightarrow{x}', t'). \end{aligned} \quad (4.8)$$

Integrating over source times t', one obtains the retarded potential

$$\phi(\overrightarrow{x}, t) = \int d^3 x' \frac{\rho \left(\overrightarrow{x}', t - \frac{|\overrightarrow{x} - \overrightarrow{x}'|}{c} \right)}{|\overrightarrow{x} - \overrightarrow{x}'|}; \quad (4.9)$$

similarly,

$$\overrightarrow{A}(\overrightarrow{x}, t) = \int d^3 x' \frac{\overrightarrow{j} \left(\overrightarrow{x}', t - \frac{|\overrightarrow{x} - \overrightarrow{x}'|}{c} \right)}{|\overrightarrow{x} - \overrightarrow{x}'|}. \quad (4.10)$$

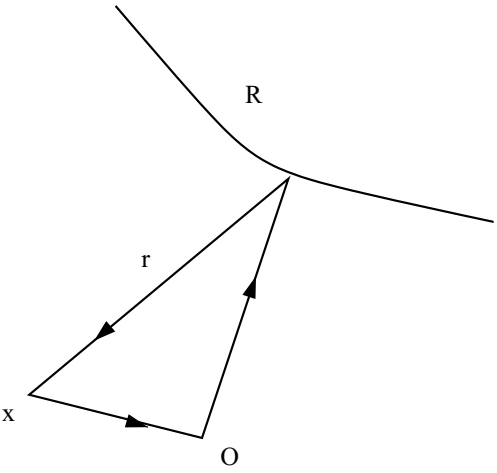


Fig. 4.1 Potentials of a moving point charge, which moves along a trajectory $\boldsymbol{R}(\tau)$ referring to the origin O. The observation point is $\boldsymbol{x}$, while $\boldsymbol{r}$ indicates the path of the radiation emitted at time τ toward the observer. That radiation will be detected in $\boldsymbol{x}$ at time $t = \tau + \frac{r}{c}$

4.3 Lienard–Wiechert Potentials

The use of the Dirac δ helps considerably in finding the potentials at the observation point $\overrightarrow{x}$ due to a point charge performing an arbitrarily assigned trajectory $\overrightarrow{R}(t)$. In the case of ϕ, from (4.8) with

$$\rho(\overrightarrow{x}', t') = e\delta(\overrightarrow{x}' - \overrightarrow{R}(t')),$$

one gets, integrating on $\overrightarrow{x}'$ first and writing τ in place of t',

$$\phi(\overrightarrow{x}, t) = e \int d\tau \frac{\delta\left(\tau - t + \frac{r(\tau)}{c}\right)}{r(\tau)}, \quad \overrightarrow{r}(\tau) = \overrightarrow{x} - \overrightarrow{R}(\tau).$$

We see that the potential is still of the Coulomb form, but is retarded: the light is received at time t, which is the emission time τ plus the propagation time $\frac{r(\tau)}{c}$; the latter depends on the position of the charge when the light is emitted.

Let τ^* denote the solution to $\tau^* = t - \frac{r(\tau^*)}{c}$.

Then,

$$\delta\left(\tau - t + \frac{r(\tau)}{c}\right) = \frac{\delta(\tau - \tau^*)}{\left|\frac{d}{d\tau}\left[\tau - t + \frac{r(\tau)}{c}\right]\right|} = \frac{\delta(\tau - \tau^*)}{\left|1 + \frac{1}{c}\frac{dr}{d\tau}\right|}.$$

Here, the absolute value is unnecessary, $|\frac{dr}{d\tau}| < c$. Since the charge is at $\overrightarrow{R}$, its velocity is $\overrightarrow{v} = \frac{d\overrightarrow{R}}{d\tau}$, while the time derivative of the distance covered by the light is

$$\frac{dr(\tau)}{d\tau} = \frac{d}{d\tau}\sqrt{\overrightarrow{r}(\tau)\cdot\overrightarrow{r}(\tau)} = \frac{1}{2\sqrt{\overrightarrow{r}\cdot\overrightarrow{r}}}2\overrightarrow{r}\cdot\frac{d\overrightarrow{r}}{d\tau};$$

since $\frac{d\overrightarrow{r}}{d\tau} = -\overrightarrow{v}$,

$$\frac{dr(\tau)}{d\tau} = -\frac{\overrightarrow{r}\cdot\overrightarrow{v}}{r}.$$

Consequently,

$$\delta\left(\tau - t + \frac{r(\tau)}{c}\right) = \frac{\delta(\tau-\tau^*)}{1-\frac{\overrightarrow{r}\cdot\overrightarrow{v}}{rc}},$$

and the potential is

$$\phi(\overrightarrow{x},t) = \frac{e}{[r-\frac{\overrightarrow{r}\cdot\overrightarrow{v}}{c}]_{\tau^*}}. \tag{4.11}$$

Starting from

$$\overrightarrow{j} = e\overrightarrow{v}\,\delta(\overrightarrow{x}-\overrightarrow{R}),$$

the same reasoning leads to the vector potential

$$\overrightarrow{A}(\overrightarrow{x},t) = \frac{e}{c}\left[\frac{\overrightarrow{v}}{r-\frac{\overrightarrow{r}\cdot\overrightarrow{v}}{c}}\right]_{\tau^*}. \tag{4.12}$$

4.4 Geometrical Optics and Fermat's Principle

Consider the propagation in a medium in which the velocity of light $v = \frac{c}{n}$ may differ from c and depend on the position due to a spatially varying refraction index $n(x, y, z)$; in addition, we may have complicated boundary conditions, like in an optical instrument. Let $\phi = u(x, y, z)e^{-i\omega t}$, with $\omega = 2\pi\nu$, represent any component of the potential or of the field of a monochromatic electromagnetic wave with frequency ν in some region of space. We wish to find an approximate solution valid at high frequencies and short wavelengths. In the wave equation $(\nabla^2 - \frac{1}{v^2}\frac{\partial^2}{\partial t^2})\phi = 0$, we may set

$$u(x, y, z) = \exp[i\nu S(x, y, z)]. \tag{4.13}$$

The new unknown S is known as the Eikonal. The wave equation becomes nonlinear:

$$\frac{4\pi^2\nu^2}{v^2} = \nu^2(\nabla S)^2 - \frac{i\omega}{2\pi}\nabla^2 S = \nu^2(\nabla S)^2 - i\nu\nabla^2 S. \tag{4.14}$$

A nonlinear equation is actually harder, but here we can simplify it since at high frequencies the last term is negligible and we are left with the Eikonal equation

$$(gradS)^2 = \frac{4\pi^2}{v^2}. \tag{4.15}$$

To explore the meaning of S, we observe that in one dimension, this becomes $\frac{dS}{dx} = \pm\frac{c}{v}$; integrating between x_A and x_B, one finds $S(x_B) - S(x_A) = \pm c\int_{x_A}^{x_B}\frac{dx}{v(x)}$. Apart from a factor $\pm c$, this is just the time the light takes to go directly from x_A to x_B. Even in 3 dimensions, the Eikonal has the same meaning, with the additional specification that the path allows for the shortest possible time. Indeed, Eq. (4.15) has the same structure as the Hamilton–Jacobi equation (2.83) where S is the action, $grad(S) = \boldsymbol{p}$ is the momentum of the material point, and E is the energy. For a free particle, it reads $(\frac{\partial S}{\partial x})^2 + (\frac{\partial S}{\partial y})^2 + (\frac{\partial S}{\partial z})^2 = E$. The equation describes corpuscles that go along trajectories that make the action an extremum. These may be interpreted as the *light particles*[3] whose trajectories are the light rays. The wavelengths may be taken to be short in a given problem when all other lengths are much longer. In this limit, there is no diffraction and the laws of Geometrical Optics are a good approximation. In the mechanical problem, S is the action and the path taken by the particle makes S an extremum. In the optical problem, S is an Eikonal, and the optical path grants the minimum time the light can take to go from A to B. The discovery that the light rays follow the quickest path is Fermat's principle, which is the oldest variational principle in Physics.[4] Consider a light ray from a point $P = (-1, y_1)$ to a point $Q = (1, y_2)$. If P is in a medium with refraction index n_1 and $Q = (1, y_2)$ is in a solid with refraction index n_2, the ray follows Snell's law[5]

$$n_1 sin(\theta_1) = n_2 sin(\theta_2).$$

The time the light takes to go from P to Q is given by $ct = n_1\sqrt{1 - (y_1 - y_S)^2} + n_2\sqrt{1 - (y_2 - y_S)^2}$; looking for the minimum as a function of y_S, we differentiate and find $n_1\frac{y_1 - y_S}{\sqrt{1-(y_1-y_S)^2}} + n_2\frac{y_2 - y_S}{\sqrt{1-(y_2-y_S)^2}} = 0$ which can be rewritten as Snell's law. A light ray in an inhomogeneous medium can be thought of as crossing a large number of interfaces following Snell's law, thereby minimizing the time it takes (Fig. 4.2).

[3] The quantum mechanical photons are emitted and adsorbed like particles but travel like waves, as we shall see below. However, the description in terms of point particles works fine at short wavelengths.

[4] Pierre de Fermat stated it in 1662.

[5] This law is actually credited to the Arab mathematician Ibn Sahl in a manuscript of 984.

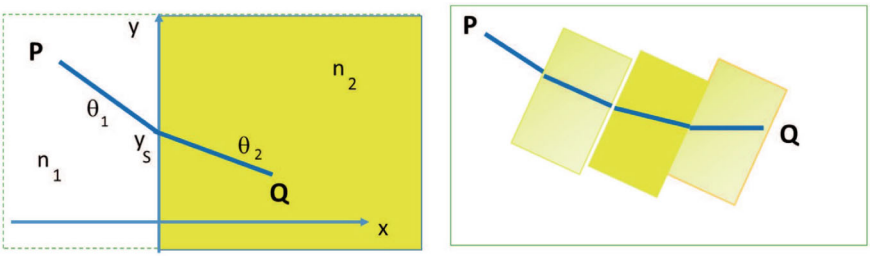


Fig. 4.2 Fermat's principle. Left panel: a light ray from a medium with refraction index n_1 to a solid with refraction index n_2. Right panel: an inhomogeneous medium behaves like a series of thin interfaces

4.5 Coherent Light

A plane wave with an electric field

$$\boldsymbol{E}(\boldsymbol{r}, t) = \boldsymbol{E_0} \exp(i\boldsymbol{k}\boldsymbol{r} - i\omega t), \tag{4.16}$$

with constant $\boldsymbol{E}_0$, is perfectly coherent light,[6] meaning that Eq. (9.3) holds everywhere; two coherent waves with different $\boldsymbol{E}_0$ and the same ω interfere. Instead, ordinary thermal light is very incoherent, and interference fringes are hardly visible. Laser light is very coherent; the degree of coherence can be measured by means of various kinds of correlation functions, like the first-order correlation function

$$g^{(1)}(\boldsymbol{r}, \tau) = \frac{\langle \boldsymbol{E}(\boldsymbol{r}, t) \cdot \boldsymbol{E}(\boldsymbol{r}, t + \tau) \rangle}{\sqrt{\langle |\boldsymbol{E}(\boldsymbol{r}, t)|^2 \rangle}}, \tag{4.17}$$

where the angular brackets stand for an average over t. The intensity correlations are also of interest, and can be made quantitative by introducing second-order correlations like

$$g^{(2)}(\boldsymbol{r}, \tau) = \frac{\langle I(\boldsymbol{r}, t) \cdot I(\boldsymbol{r}, t + \tau) \rangle}{\sqrt{\langle |I(\boldsymbol{r}, t)|^2 \rangle}}. \tag{4.18}$$

4.5.1 The Measurement of Stellar Diameters: Hanbury Brown and Twiss Effect (Classical Case)

Disappointingly, even nearby stars appear point-like in the most powerful telescopes. The Very Large Telescope under favorable conditions attains a resolution of 2×10^{-3} arc seconds; Sirius has a diameter 2.617×10^6 km and its distance obtained by

[6] We shall have more to say about coherent states in Chap. 16 when dealing with Photons and in Chap. 25 when dealing with Lasers.

parallax is 8.6 light years. A light year is about $9.46\ 10^{12}$ km. This implies an angular diameter of about 5×10^{-6} radiants, which is below the detection limit. However, the diameter quoted above was obtained, thanks an to interesting interferometric technique introduced in 1954 by Hanbury Brown and Twiss. It is based on a rather surprising concept known as Hanbury Brown and Twiss effect. Their method uses two detectors D1 and D2 on the ground separated by a distance of the order of meters, producing photocurrents $I_1(t)$ and $I_2(t)$ that go to an electronic device called a correlator. The correlator produces the output

$$C = \frac{\langle I_1(t) I_2(t) \rangle}{\langle I_1(t) \rangle \langle I_2(t) \rangle}. \tag{4.19}$$

The time average intensities $\bar{I}$ at the two detectors are the same, but the relevant information is contained in the fluctuations. Now, C is again a correlation very similar to the one in Eq. (4.18). If $I_1(t)$ is proportional to a square electric field $E(t)^2$ and $I_2(t)$ is proportional to a square electric field $E(t-\tau)^2$, it is clear that C is suitable to measure the delay τ. In the observation of stars there are two sources of delay. One is more obvious, and is due to the fact that in general the two detectors are at different distances from the star. The second is due to the fact that the star has a finite size, of the order of a million Km, so the waves coming from different points of the photosphere do not reach both detectors simultaneously. Thus, an analysis of C allows to measure the diameter. A practical limitation of this method arises from the atmospheric turbulence that causes rapid fluctuations of the phase of the electric field. Recent Quantum versions of the Hanbury Brown and Twiss effect are deferred to Sect. 25.7.

4.6 The Hall Effect

This section is devoted to the Hall effect, as it was known before the discovery of the quantum Hall effects in 1980 and subsequent years. Edwin Hall in his 1879 experiment aimed to measure the effect of an applied magnetic field on the current flowing in a conductor. Note that he conceived the effect many years before the discovery of the electron by Thomson in 1897; in 1879 physicists knew how to excite currents but had no microscopic understanding of their nature. The theory was formulated much later through several developments. The classical Drude theory which was formulated in the early twentieth century is outlined below. The quantum theory is outlined in Chaps. 22 and 26. The Quantum Hall effect which is an extra development discovered in 1980 is deferred to Chap. 26.

We cut out a rectangle from a metal plate, connect the opposite sides A and B to an electrical circuit and switch a magnetic field $\overrightarrow{B}$ orthogonal to the metal plate and an electric field E_x between A and B; we let the xy plane coincide with the upper surface of the metal and the z axis orthogonal to it and parallel (or antiparallel) to

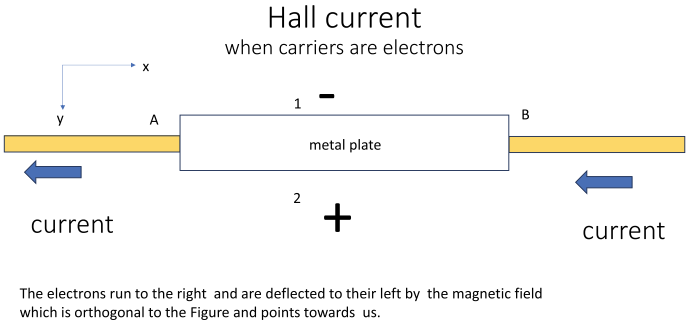


Fig. 4.3 The geometry of the Hall experiment. The electrons run to the right and are deflected to their left by the magnetic field, which is orthogonal to the plate points up toward us

$\vec{B}$. So, the current flowing in the circuit is J_x. Finally, we measure the electric field $\vec{E}_y$ across the plate (see figure), which is due to the fact that the Lorentz force acting on the charge carriers produces a transverse polarization of the sample (Figs. 4.2, 4.3 and 4.4).

In this way, Hall determined that the charge carriers were negative, many years before the discovery of the electron by Thompson in 1897. Strikingly, some semiconductors have positive carriers, and this remained unexplained until Quantum Theory of Solids clarified that the carriers are holes.[7] The Hall coefficient is defined as

$$R_H = \frac{E_y}{J_x H}. \tag{4.20}$$

At the beginning of the new century, Drude[8] proposed a classical model which is surprisingly successful (considering its simplicity) and among others allows an approximate calculation of R_H. The equation of motion of the electron is

$$\frac{d\vec{p}}{dt} = -e(\vec{E} + \frac{\vec{p}}{mc} \wedge \vec{B}) - \frac{\vec{p}}{\tau}; \tag{4.21}$$

here the first term on the R.H.S. is the familiar Lorentz force, while the last term is the contribution of a phenomenological friction force which represents the effect of the collisions causing a damping of the electron. In steady state conditions, $\frac{d\vec{p}}{dt}$ is supposed to vanish. Thus, one finds

$$\sigma_0 E_x = \tau\omega_c J_y + j_x, \sigma_0 E_y = -\tau\omega_c J_x + j_y, \tag{4.22}$$

[7] Holes are empty electron states in the valence band of semiconductors (see Chaps. 22 and 25 for some concepts on the quantum theory of solids).

[8] Paul Karl Ludvig Drude (1863–1906) German Physicist.

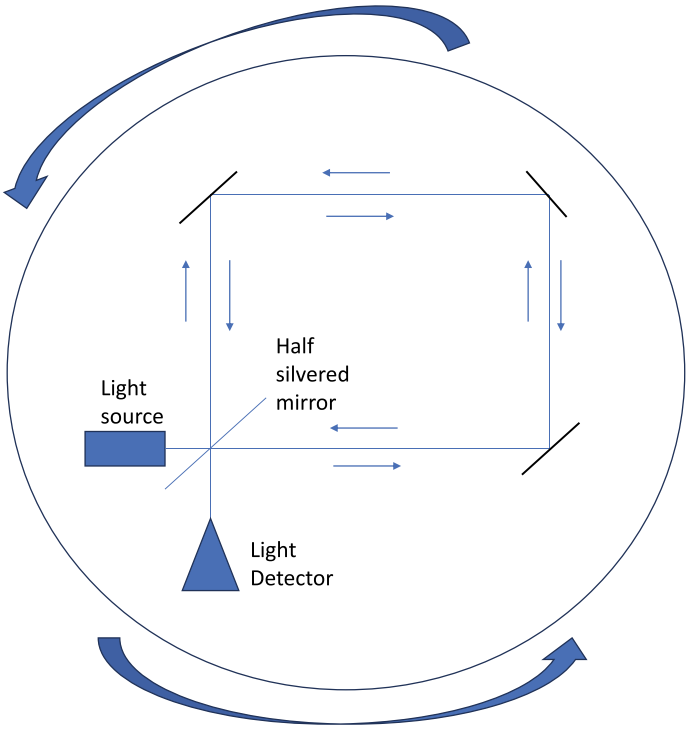


Fig. 4.4 Sagnac interferometer. A light beam from the source is split by a half silvered mirror in two beams that follow opposite paths before ending in the detector. The entire apparatus is mounted on a support that is rotated, and this creates a phase shift that the detector measures

where

$$\vec{j} = ne\vec{v}, \tag{4.23}$$

n is the number density of the charge carriers,

$$\sigma_0 = \frac{ne^2\tau}{m} \tag{4.24}$$

is the Drude conductivity, and $J_y = 0$ in the present geometry. Finally,

$$E_y = -\frac{Hj_x}{nec} \tag{4.25}$$

and

$$R_H = -\frac{1}{nec}. \tag{4.26}$$

The $\frac{1}{n}$ dependence trivially stems from the fact that one measures the electric potential and the force is proportional to n. The Hall effect is used to measure the sign and magnitude of the carrier density n. The Hall resistance defined as

$$R_H = \frac{E_y}{J_x} \tag{4.27}$$

is predicted to be proportional to the magnetic field H. From the classical viewpoint this is logical, since E_y must balance the Lorentz force. This simple theory is in rather good agreement with experiments on metals.[9] For example, for Li, the experimental $R_H = -13.2$ is close to the Drude result $R_H = -12.8$, and a similar agreement is obtained for Alkali metals, while for other metals the comparison with an experiment is much worse (this is typical for other properties, like the optical properties, in the Drude model) but the sign is predicted correctly in most cases. One concludes that the in the Drude theory, the Hall coefficient measures the sign and the density of carriers; the information on the density is related to the plasma frequency (see below). The methods of Quantum Transport (see Chap. 26) allow a more satisfactory treatment of the Hall effect. However, the term "Quantum Hall effect" conventionally refers to related but different phenomena which were discovered in 1980 and following years, which require two-dimensional samples (that is, thin films) in very strong magnetic fields/several Tesla and very low temperatures (see Chap. 26).

4.7 Sagnac Effect

Georges Sagnac published in 1913 the results of an experiment showing that the speed of light is independent of the velocity of the source.

Consider a circular loop of radius R with two light pulses sent symmetrically, one in clockwise and the other in counterclockwise direction, while the loop rotates clockwise with angular frequency ω. The rotation breaks the symmetry and causes a delay

$$\Delta t = 2\pi R\left(\frac{1}{c-v} - \frac{1}{c+v}\right) = \frac{4A\omega}{c^2 - v^2} \tag{4.28}$$

where A is the area of the loop, for the clockwise pulse relative to the other. One can prove that nothing changes if the loop is not circular. Such a device was used by Michelson and Gale in 1925 to measure the absolute rotation of the Earth. Devices based on the same principles but operated using laser light are used by aircraft, spacecraft and the like, in order to detect very small rotation rates.

Sagnac did not agree with Einstein's relativity at the time; however there is no conflict with it. In Chap. 8, we shall describe the operation from the standpoint of General Relativity.

[9] See Thomas DeMastri, Arxiv cond-mat December 2021.

He proposed the experiment as a way to demonstrate his ether model and explained the effect without using relativity theory.

The effect has been observed experimentally in a wide range of wavelengths, from radio to X-rays and is of considerable practical importance in Global Positioning System.

The Sagnac effect is used in laser gyroscopes, which are instruments used to measure and maintain orientation, particularly for inertial navigation systems, like aircraft, and traditionally were mechanical (based on angular momentum conservation of rotating masses).

Chapter 5
Thermal Physics

Any macroscopic object in thermal equilibrium contains so many particles in chaotic motion that the methods of Mechanics are totally useless. This chapter is a simple introduction to Thermodynamics and classical Statistical Mechanics.

5.1 The Principles of Thermodynamics

Thermodynamics is an axiomatic part of Theoretical Physics, which is presented as a set of phenomenological axioms or principles. Any investigation into the reasons why the principles are true and how they are related to micro-Physics belongs to the domain of Statistical Mechanics, while Thermodynamics gives guidelines that are never contradicted by experiments, at least for macroscopic objects. Actually, the principles are simple and part of the common wisdom by now, but Thermodynamics deduces profound consequences.

Zeroth Principle

The *zeroth principle of Thermodynamics* states that two bodies can be in thermal equilibrium and that such a state is transitive: if bodies A and B are in equilibrium and B is in equilibrium with C, then A is in equilibrium with C.

This allows us to introduce the concept of temperature, which has a precise meaning only in equilibrium.

First Principle

Nowadays, the invoice of electricity reminds us that heat requires energy, but the physical nature of heat has long remained unknown. It was really a scoop when James Prescott Joule in the 1840s measured the mechanical equivalent of the calorie,

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_5

which is the amount of heat required to raise the temperature of 1 g of water by 1 °C at the pressure of 1 atmosphere; the modern value is 4.1855 J. The *First Principle of Thermodynamics* establishes the conservation of mechanical and thermal energy. If some amount δQ of heat is supplied to any system, the effect is to increase its internal energy and/or to cause some work to be performed. So,

$$\delta Q = dU + PdV, \tag{5.1}$$

where dU is the increase in internal energy, P the pressure, V the volume, and PdV is the work performed by the system. This result implies that a *perpetual motion machine of the first kind (producing work without a power supply)* cannot exist.

Second Principle

However, it is possible to make a motor, that is, a thermal machine that converts thermal energy into mechanical energy. To proceed, we need the notion of a *reversible* transformation. Such a process must take place so slowly and gradually that at all times, the system has pressure, temperature and volume very close to the equilibrium values; therefore, it must be possible to do the same transformation in reverse, driving the system back to the same states of quasi-equilibrium.[1]

Aiming to build a motor, we need first to build a cycle,, i.e., a process in which the fluid after producing some work, returns to its initial state ($\Delta U = 0$), ready to iterate the process. Consider a fluid such that the state is characterized by (P, V, T), that is, pressure, volume and temperature. Carnot's cycle is defined by the reversible transformations in Fig. 5.1, namely two isotherms at temperatures T_1 and $T_2 < T_1$ and two adiabatics (no heat exchanged). The machine is reversible and can work as a motor or a refrigerator. In Fig. 5.1, we see its use as a thermal machine, which

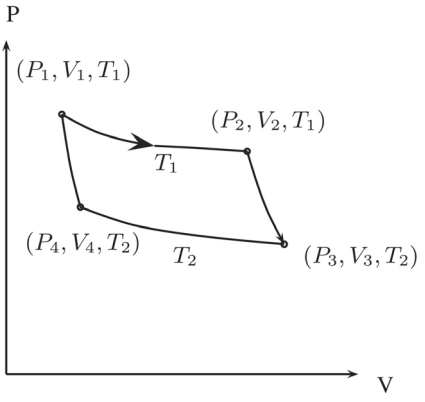


Fig. 5.1 Carnot's cycle

[1] A piece of diamond is out of equilibrium under normal conditions, because Graphite is the equilibrium form of Carbon; however since the kinetics of the transformation is extremely slow, one can do reversible transformations on diamond as well. Actually, the seemingly simple concepts on which Thermodynamics is based are far from obvious.

absorbs heat Q_1 at temperature T_1 and loses heat Q_2 at a lower temperature T_2. The first principle grants that the work done in the reversible cycle is

$$W = \oint PdV = Q_1 - Q_2; \tag{5.2}$$

the efficiency is defined as

$$\eta = \frac{W}{Q_1} = 1 - \frac{Q_2}{Q_1}.$$

The possibility of trading heat for work implies that the heat content and the work done by a fluid depend on the previous history of the machine. In mathematical terms, δQ is not the differential of a function of the parameters that specify the state, since in a cycle, $\Delta Q \equiv \oint \delta Q$ is non-vanishing, and the fluid can have undergone many cycles before returning to a given state.

The first principle would allow for $\eta = 1$ (total conversion of the heat from a source to work, equivalent to unlimited energy for free).

The situation changes radically when we introduce the Second Principle. The *Second Principle* (stated by Kelvin and Planck) says that it is impossible to build a *perpetual motion machine of the second kind*. This would consist of a device that makes mechanical work at the expense of the heat taken from one source. Clausius states the same principle in an even simpler and more striking way: heat goes spontaneously from hot to cold bodies, and work is needed to force it to go from cold to hot. Actually, the two statements are equivalent. If it were possible to extract work from the cold source (in violation of the Kelvin–Planck statement), then we could use the work to heat the hot source, thus violating the Clausius statement. If it were possible to violate the Clausius statement, we could produce and increase the temperature difference between two bodies at the expense of the energy of the cold body; then we could action a Carnot cycle and obtain work, still at the expense of the cold body, violating the Kelvin–Planck statement.

The second principle puts severe limits on what can be done by the Carnot cycle. The efficiency of a Carnot cycle depends only on the temperature of the heat sources. Indeed, if the Carnot cycles C_1 and C_2 between the same temperatures had efficiencies η_1 and η_2 with $\eta_1 > \eta_2$, we could use the first cycle to fool the second principle in the second cycle.

Problem 5.1 How?

Solution 5.1 C_1 would be used as a motor to produce the amount W_1 of work. Using work $W_2 < W_1$, one could use C_2 as a freezer, giving back to the hot source the heat taken by C_1. We should be left with an amount $W_1 - W_2$ of work, produced at the expenses of the cold source.

For the same reason, no other cycle can beat Carnot's cycle in efficiency.

Now we are in a position to define the thermodynamic temperature, which is one of the basic achievements of thermal Physics. It is clear that any definition

based on Mercury, ideal gases or any other (real or ideal) material can be used in a limited interval of temperatures and has no universal significance. But since η depends only on the temperatures, one can use measurements of η to define a universal thermometric scale. More precisely, we can define the absolute temperature T such that in a Carnot cycle,

$$\frac{T_2}{T_1} = \frac{Q_2}{Q_1}; \tag{5.3}$$

this law fixes all temperatures, including the extreme ones, in terms of a referential one, without having to rely on the properties of any substance. Now, the efficiency becomes

$$\eta = \frac{W}{Q_1} = 1 - \frac{T_2}{T_1}.$$

Now we see that if the heat sink is very cold, the efficiency approaches 1; however T_2 can never be negative, because otherwise we should be extracting limitless energy from the cold source. Thus, an absolute zero must exist.

We need to be able to measure amounts of heat. Having done that, we have the extra bonus that in any cycle (and in any combination of cycles),

$$\frac{Q_2}{T_2} - \frac{Q_1}{T_1} = 0 = \oint \frac{\delta Q}{T}.$$

Thus, the *second principle of Thermodynamics* grants that in any reversible transformation at temperature T,

$$\frac{\delta Q}{T} = dS \tag{5.4}$$

is the exact differential of a function of temperature, pressure and any other relevant parameter, which is called Entropy. In any reversible cycle,

$$\oint dS = 0. \tag{5.5}$$

The temperature is dimensionless, while S is energy.

5.2 Thermodynamic Potentials

Putting together the first and the second principles, (5.1) and (5.5), we find that two equilibrium states close to each other are related by

$$TdS = dU + PdV. \tag{5.6}$$

No doubt, the internal energy U is a function of state, with a well-defined value independent of the history; this result shows that it is precisely $U = U(S, V)$, that is, it depends on S e V. However, it is not easy to work with it, since S is extensive (i.e., proportional to the size of the system) like U. As a result, it is not so easy to measure: there is no such instrument as a *entropometer*.

From (5.6), we obtain the temperature,

$$T = \left(\frac{\partial U}{\partial S}\right)_V \tag{5.7}$$

(derivative taken at constant volume V); T can be measured easily with a thermometer. To change the independent variable, we introduce chemical potentials by means of Legendre transformations. The free energy is suitable for us

$$F = U - TS. \tag{5.8}$$

The differential is, using (5.6),

$$dF = dU - TdS - SdT = -PdV - SdT. \tag{5.9}$$

Hence,

$$\left(\frac{\partial F}{\partial V}\right)_T = -P, \ \left(\frac{\partial F}{\partial T}\right)_V = -S. \tag{5.10}$$

$F = F(V, T)$ is particularly well-suited for processes at constant volume, because only the T dependence is left. However, keeping V constant during a transformation may be difficult. Usually, one prefers to let V vary while keeping the pressure P constant. A possible choice is the enthalpy $H(S, P)$ defined by

$$H \equiv U + PV. \tag{5.11}$$

One finds that

$$dH = TdS + VdP. \tag{5.12}$$

This contains the differential of entropy, which has the drawbacks we have just discussed. The natural choice for most experimental conditions is the Gibbs free energy

$$G = U + PV - TS. \tag{5.13}$$

Its differential reads as

$$dG = -SdT + VdP. \tag{5.14}$$

5.2.1 Equation of Clausius–Clapeyron

Many substances have two phases with a border in the P-T (Pressure–Temperature) diagram. The Clausius–Clapeyron equation relates the volume change and the latent heat of the transition with the shape of the boundary. Since P and T vary along the boundary, the appropriate potential to consider is G, starting from the fact that along the boundary between the phases A and B, the equilibrium condition is $G_A = G_B$. Therefore the phase boundary must be such that $dG_A = dG_B$:

$$\left(\frac{\partial G_A}{\partial T}\right) dT + \left(\frac{\partial G_A}{\partial p}\right) dP = \left(\frac{\partial G_B}{\partial T}\right) dT + \left(\frac{\partial G_B}{\partial p}\right) dP. \tag{5.15}$$

This implies

$$V_A dP - S_A dT = V_B dP - S_B dT \tag{5.16}$$

$$\frac{dP}{dT} = \frac{S_B - S_A}{V_B - V_A}. \tag{5.17}$$

Hence,

$$S_B - S_A = \int_A^B \frac{\delta Q}{T} = \frac{1}{T} \int_A^B dQ = \frac{\lambda}{T}, \tag{5.18}$$

and, finally,

$$\frac{dP}{dT} = \frac{\lambda}{T(V_B - V_A)}. \tag{5.19}$$

5.2.2 Maxwell Relations

Comparing (5.6) with

$$dU = \left(\frac{\partial U}{\partial S}\right)_V dS + \left(\frac{\partial U}{\partial V}\right)_S dV,$$

one finds that

$$T = \left(\frac{\partial U}{\partial S}\right)_V, \quad P = -\left(\frac{\partial U}{\partial V}\right)_S. \tag{5.20}$$

Since the mixed derivatives of $U(S, V)$ must agree, one obtains the Maxwell relation

$$\left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial P}{\partial S}\right)_V. \tag{5.21}$$

It is clear that the left-hand side is easier to measure experimentally than the right-hand side. Other Maxwell relations can be obtained from the other potentials. From the differential of Enthalpy, one finds

$$T = \left(\frac{\partial H}{\partial S}\right)_P, \quad V = -\left(\frac{\partial H}{\partial P}\right)_S. \tag{5.22}$$

Problem 5.2 What Maxwell relation follows from Eq. (5.22)?

Solution 5.2

$$\left(\frac{\partial T}{\partial P}\right)_S = -\left(\frac{\partial V}{\partial S}\right)_P. \tag{5.23}$$

Starting from the differential of F, one finds that

$$-P = \left(\frac{\partial F}{\partial V}\right)_T, \quad -S = \left(\frac{\partial F}{\partial T}\right)_V, \tag{5.24}$$

and one obtains the Maxwell relation

$$\left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial P}{\partial T}\right)_S. \tag{5.25}$$

Starting from the differential of G, one finds that

$$-S = \left(\frac{\partial G}{\partial T}\right)_P, \quad V = \left(\frac{\partial G}{\partial P}\right)_T, \tag{5.26}$$

and one obtains the Maxwell relation

$$-\left(\frac{\partial S}{\partial P}\right)_T = \left(\frac{\partial V}{\partial T}\right)_P. \tag{5.27}$$

5.2.3 Chemical Potential

Often, it is preferable to work with a thermodynamic system embedded in an environment (or thermal bath) with which it can exchange not only heat but particles, too. When the number of particles is variable, one introduces the chemical potential μ by extending (5.6):

$$dU = TdS - PdV + \mu dN.$$

This is familiar, since we ground our electrical apparatus to avoid shock hazards; the electric potential is the μ that we must control.

5.2.4 Perfect Classical Gas

For a perfect gas, the thermodynamic properties depend on the number f of degrees of freedom of the molecules; $f = 3$ for monoatomic gases to which only translation contributes, but diatomic molecules have $f = 5$ due to rotational freedom; extra degrees of freedom at relatively high temperatures (several hundreds of Kelvin degrees) result from molecular vibrations. The internal energy is

$$U = \frac{f}{2} N K_B T. \tag{5.28}$$

The equation of state is the Clapeyron law

$$PV = NK_B T = nRT, \tag{5.29}$$

where $K_B = 1.381 \times 10^{-16} \frac{\text{erg}}{^\circ\text{K}}$ is Boltzmann's constant, n the number of moles, and $R = N_A K_B$ is called the gas constant. For an adiabatic expansion ($\delta Q = 0$), inserting (5.28) into $dU + PdV = 0$, one finds

$$(1 + \frac{f}{2}) PdV + \frac{f}{2} VdP = 0.$$

Integrating, one finds PV^γ =constant, where

$$\gamma = \frac{f+2}{f}.$$

The specific heat is by definition the thermal capacity of a mole ($N = N_A$). The specific heat at constant volume is

$$C_V = \left(\frac{\delta Q}{dT}\right)_V. \tag{5.30}$$

One obtains

$$C_V = T \left(\frac{\partial S}{dT}\right)_V. \tag{5.31}$$

The specific heat at constant pressure, which is generally easier to measure, is

$$C_P = T \left(\frac{dS}{dT}\right)_P.$$

Problem 5.3 Calculate $\frac{C_P}{C_V}$ for a perfect gas.

Solution 5.3 One finds $C_V = \frac{f}{2} N_A K_B$, $C_P = (1 + \frac{f}{2}) N_A K_B$. At constant volume, the term PdV is discarded, while at constant pressure, V can be eliminated using the Clapeyron equation. From this result, one finds that

$$\frac{C_P}{C_V} = \gamma. \tag{5.32}$$

5.2.5 *The Arrow of Time*

The entropy is maximum in equilibrium. Take a container A having volume V_A that is part of a larger container B having volume V_B; A is separated from the rest of B by a partition and is filled with some perfect gas. Suppose that the walls are thermal insulators. When the gas is in equilibrium at temperature T_A, we suddenly remove the partition. I think that you will easily agree that the gas will occupy all the volume B, and after some time, it will be in equilibrium again at some temperature T_B.

Actually, since in the process, $\int \delta Q = 0$ and the work done on the gas or by the gas is nothing, $\Delta U = 0$. For a perfect gas, $U = U(T)$, so $T_B = T_A$. The process is a typical *irreversible process*, and the second principle implies that

$$S_B - S_A > \int_A^B \frac{\delta Q}{T} = 0.$$

To compute $S_B - S_A$, consider a reversible isothermal expansion. Since $U = U(T)$, the expansion is isothermal if $dU = 0$. To expand the gas, we slowly move a piston, but the gas makes work on the piston when the volume increases by dV. Now, $dU = 0 \rightarrow \delta Q = PdV$, and we must supply this heat in the process. The change in entropy is

$$S_B - S_A = \int_A^B \frac{PdV}{T},$$

and inserting P from (5.29), one obtains

$$S_B - S_A = N K_B \log\left(\frac{V_B}{V_A}\right). \tag{5.33}$$

The motion of each molecule is mechanically reversible, but the time-reversed process never occurs. The future can be distinguished from the past because of the sign of $S_B - S_A$. The spontaneous evolution of any system increases the entropy, therefore the entropy is maximum at equilibrium. This is the *thermodynamic arrow of time*.

More generally, we can consider a reversible transformation in which, besides changing the volume we also exchange heat with n moles of perfect gas. This can be

done by adding an adiabatic transformation. Then, $N = nN_A$ and $dS = nC_V dT + PdV$, and inserting Clapeyron's law,

$$dS = nC_V dT + nR\frac{dV}{V};$$

integrating,

$$S_B - S_A = nC_V \ln\left(\frac{T_B}{T_A}\right) + nR \ln\left(\frac{V_B}{V_A}\right). \tag{5.34}$$

It is important to note that the entropy is an extensive quantity, that is, each molecule contributes a molecular entropy $\frac{S_B - S_A}{N}$, and the set of two independent tanks of gas has an entropy that is the sum of the two entropies. This is like the kinetic energy of molecules that sum up to give the internal energy of the gas; however, the observation is more subtle, since the molecular entropy depends on T, which is not a property of the single molecule.

The Universe was hot initially, yet, the entropy of the Universe is Growing (how this comes about is a problem of Cosmology). Even the Universe as a whole (Sect. 8.12) is growing old, the star formation rate is decreasing and the synthesis of heavy elements inside stars is clearly an irreversible process. The cosmological models that proposes cyclic or pulsating universes must face the basic problem of "resetting" the entropy at the start of each cosmic cycle.

Today, CPT is considered a fundamental symmetry of Nature, with C the charge conjugation connecting particles and antiparticles, P the parity and T the time reversal. The (unexpected) violation of parity by the weak interactions was discovered in 1956 by Chien-Schiung Wu in the β decay of Co60; James Cronin and Val Fitch unexpectedly discovered in 1964 that the decays of neutral kaons in two and three pions show a very weak violation of CP, which implies a very weak violation of T. This is the most mysterious arrow of time.

5.2.6 *Low Temperatures and the Third Principle*

The *third principle of Thermodynamics* was formulated by Walther Nernst in 1906–1912. It states that $S \to 0$ for $T \to 0$. This removes the arbitrariness that was left by the second principle. One can show that this requires that the specific heat C of every perfect crystal tends to 0 at absolute zero, provided that the lowest state of the system is unique. More generally, for any system $\frac{\delta Q}{T} = \frac{C\delta T}{T}$ must stay small for any process with $T \to 0$, making it increasingly difficult to extract heat from a substance when the temperature gets lower. The most important implication is that it is impossible to reach the absolute zero in a finite number of operations. Today, there are people working with ultra-cold atoms at micro-Kelvin temperatures obtained by laser cooling techniques and manipulating them to make studies on the quantum behavior of matter under such conditions. The 1997 Nobel Prize in Physics was given

to Steven Chu, Claude Cohen-Tannoudij and William D. Phillips for the development of methods to cool and trap atoms with laser light. One could wonder what is the significance of the third principle when a disordered mixture is brought to ultra-low temperatures; how is it possible that the measure of disorder S gets small by cooling? But at absolute zero, the system should get ordered, and S=0 corresponds to a perfect crystal. The apparent paradox comes from the fact that the kinetics for approaching the equilibrium $S = 0$ state gets extremely slow at low T.

5.3 Black Body

The objects that we observe in everyday life are colored according of their nature. Viewed in white light, a body turns blue if it absorbs mainly red light and red if it better absorbs blue radiation; it looks white if it reflects all that is visible. A body reflects some light, even without being an ideal mirror. (A mirror also preserves direction and images.) Some bodies absorb little light and reflect little; they look transparent and colorless. It is the light that the body reflects that we see and gives it a color. If a body in daylight reflects little visible light, it appears black; but coal also reflects a bit. An ideal black body absorbs all radiation completely. Such a body is an idealization of great importance in the historical development of Theoretical Physics. It is due to Kirchhoff.[2] The starting point was a focus on equilibrium conditions. Only a deep reflection allowed for the discovery that this color variability, which is so necessary to our life, is due not only to the nature of the body but primarily to the color of light, and especially to the lack of equilibrium between radiation and matter. (Life is itself a phenomenon that is far from equilibrium.) According to the first principle of Thermodynamics, energy must be conserved. For a body in equilibrium, the energy that is absorbed is re-emitted. This leads to the idea of an equilibrium distribution in frequency of the radiation absorbed and emitted by all bodies.

The ancient philosophers did not even remotely imagine such things. Everybody knew that a piece of iron when heated starts to emit light that changes from red brown to red to orange to yellow with increasing temperature, and becomes much more bright. They also knew that all hot bodies behave in essentially the same way. Of course, the problem is complicated by the fact that our eye sees only wavelengths between 4000 and 8000 ρA. In the nineteenth century, experiments led to Wien's law $\nu_{max} \propto T$, saying that the frequency ν where maximum intensity is issued is proportional to the temperature. The power emitted increases as ν^4, as we shall see.

An ideal black body absorbs all radiation, but such a thing does not exist. A closed cavity in equilibrium at temperature T is filled with thermal radiation, which is emitted from the walls. In equilibrium, as much has to be absorbed as is emitted.

[2] Gustav Kirchhoff was a German physicist (Königsberg (now Kaliningrad, Russia), 1824–Berlin, 1887). He established the well-known laws on linear electric circuits, and, together with Bunsen, invented the spectroscope; he also discovered the elements Cesium and Rubidium. He wrote "Vorlesungen *ü*ber mathematische Physik".

Through a small hole in the walls, one can observe this radiation. This is an experimental realization of the black body, since every radiation entering the hole from outside will be partially reflected by the walls many times and eventually will be absorbed.

Kirchhoff Laws

Consider a cavity in equilibrium at temperature T. It is full of isotropic radiation in equilibrium with the walls. Let $u(\nu, T)$ denote the energy per unit volume and frequency ($[u(\nu, T)] = \frac{\text{erg.s}}{\text{cm}^3}$). The energy density is $u(T) = \int_0^\infty u(\nu, T)d\nu$. Gustav Kirchhoff, in 1859, has shown that $u(\nu, T)$ does not depend on the nature of the walls of the cavity. Suppose two different cavities A and B at the same temperature had different densities $u_A(\nu, T)$ and $u_B(\nu, T)$. We could let them exchange radiation and, using a color filter, we could allow the exchange of energy at frequencies such that $u_A(\nu, T) > u_B(\nu, T)$ but not at frequencies such that $u_A(\nu, T) < u_B(\nu, T)$. In this way, B should become hotter than A. The perpetual motion machine of the second kind would be operating in contrast to the Second Principle. Kirchhoff concluded that $u(\nu, T)$ must be some universal function. One can put objects of different colors, bright or opaque, but at equilibrium the same $u(\nu, T)$ is obtained.

To explain that, one must consider the energy balance of a surface element dS of the walls in the unit time. The energy emitted in a frequency interval $d\nu$ centered around ν is $e(\nu, T)dS$, where e is the emissive power which depends on the nature of the body. On the other hand, when isotropic radiation with frequency ν falls on the same surface element, the nature of the body also determines the absorption power $a(\nu, T)$ which is the ratio of absorbed energy over incidence energy. An ideal body with $a \equiv 1$ is called black. Remarkably, Kirchhoff found that the ratio $\frac{e(\nu,T)}{a(\nu,T)}$ does not depend on the body. Indeed, the same amount of energy in the interval $d\nu$ falls in the unit time over each dS of the walls and of the surfaces of all the bodies therein; a fraction $a(\nu, T) \cdot dE$ is absorbed, but an equal amount $e(\nu, T) \cdot dS$ must be emitted. There is a separate equilibrium condition at each frequency. For two bodies 1 and 2, the equilibrium conditions are

$$\begin{cases} e_1(\nu, T)dS = a_1(\nu, T)dE(\nu, T), \\ e_2(\nu, T)dS = a_2(\nu, T)dE(\nu, T), \end{cases} \tag{5.35}$$

and we may conclude that

$$\frac{e_1(\nu, T)}{a_1(\nu, T)} = \frac{e_2(\nu, T)}{a_2(\nu, T)} = \frac{dE(\nu, T)}{dS}. \tag{5.36}$$

For the black body, a=1. The universal function $\frac{dE(\nu,T)}{dS}$ is the *emitting power* of the black body; it is the energy that is emitted for unit time and frequency by the unit surface of the black body. Experimentally, one can make a good black body by producing a small hole on the surface of a cavity kept at temperature T. If the temperature is several hundred degrees Celsius, the hole appears to an external observer

as a bright spot. Its absorbing power is high (nearly all the incident radiation bounces on the walls many times and is absorbed). The emissive power is very high. We now prove that

$$\frac{dE(\nu, T)}{dS} = \frac{c}{4}u(\nu, T). \tag{5.37}$$

To see the relation between dE and u, let us take the surface element dS as the origin of spherical coordinates; a volume element of the cavity at (r, θ, ϕ) is

$$dV = r^2 dr d\Omega = r^2 dr d\phi \sin(\theta) d\theta;$$

at time $t = 0$, dV contains the energy $u(\nu, T)d\nu dV$ in the frequency interval $d\nu$. Such radiation is isotropic, and a fraction will arrive at dS at time $t = \frac{r}{c}$. Seen from dV, the surface element is reduced by a factor of $\cos(\theta)$ and occupies a fraction $f dS = \frac{dS\cos(\theta)}{4\pi r^2}$ of the spherical surface of radius r. In the time between $t = 0$ and $t = \frac{r}{c}$, dS receives the contributions of the elements dV up to a distance r;

$$dE = dS \int_0^{ct} dr r^2 \int_0^{2\pi} d\phi \int_0^{\frac{\pi}{2}} d\theta \sin(\theta) f,$$

the integral over θ covers the half space "above the surface". Then, we must divide by t to get the energy per unit time. The result is

$$dE(\nu, T) = c\frac{u(\nu, T)dS}{4\pi} 2\pi \int_0^1 d\cos(\theta)\cos(\theta) = \frac{c}{4}u(\nu, T)dS.$$

This is Eq. (5.37).

Stefan–Boltzmann Law

Consider an empty cavity with volume V in thermal equilibrium at temperature T. We know already that the cavity contains electromagnetic radiation, the black body radiation. According to the Kirchhoff laws, the spectral components $u(\nu, T)$ obey a universal law (which will turn out to be Planck's law); the total internal energy is

$$U = Vu(T). \tag{5.38}$$

$U = U(V, T)$ does not depend purely on T, as for the classical ideal gas, but also on volume. There is a pressure on the walls. It is known from Electromagnetism that the radiation pressure P is related to the energy density u by

$$P = \frac{u}{3}. \tag{5.39}$$

A thought experiment is needed here. The cavity is closed by a (frictionless) piston in contact with the heat source at temperature T, and the radiation is used as the fluid to make a reversible infinitesimal Carnot cycle. The four phases are

A isothermal expansion $V \to V + \Delta V$ at temperature T. The internal energy grows by $u\Delta V$ and the work $P\Delta V$ is performed; therefore, from the First Principle, we know that from the source at temperature T, an amount of heath $Q = (P + u)\Delta V$ is extracted;
B adiabatic expansion (cavity and piston are now thermally isolated) $V + \Delta V \to V + \Delta V + dV$ with infinitesimal dV. The system makes work at the expense of its internal energy; the temperature decreases to $T - dT$, while the pressure drops to $P - dP$ with $dP = \frac{dP}{dT} dT$;
C isothermal compression with the cavity in contact with the source at temperature $T - dT$. The volume decreases from $V + \Delta V + dV$ and crosses the adiabatic through the initial state. The volume decreases by ΔV (neglecting higher order corrections). The heat given to the cold source at temperature $T - dT$ is $Q = [(P - dP) + u]\,\Delta V$;
D an adiabatic compression closes the cycle.

Up to higher order infinitesimals, the work done in the cycle is $dW = dP\Delta V$; the efficiency is

$$\eta = \frac{dW}{Q} = \frac{dP\Delta V}{(P + u)\Delta V} = \frac{dP}{dT}\frac{dT}{P + u}. \tag{5.40}$$

But the Second Principle says that the efficiency must be $\frac{dT}{T}$. So,

$$\frac{dP}{dT} = \frac{P + u}{T}. \tag{5.41}$$

Substituting (5.39) and integrating, one finds

$$u = aT^4. \tag{5.42}$$

This is the *Stefan–Boltzmann* law. A hot black body is in equilibrium with an energy density, which grows with the fourth power of temperature; this is also the trend of the power radiated by a black body in the absence of equilibrium, when the body cools down. The stars have internal heat sources and emit by irradiation in agreement with the *Stefan–Boltzmann* law, since they are black to a good approximation. Stefan obtained the result experimentally. The theory was formulated by Boltzmann[3] in 1884; the constant a remained an empirical parameter for a long time, until it was determined with the advent of Quantum Mechanics.

Problem 5.4 The black body radiation carries entropy S. Consider the phases A and B of the above Carnot cycle; calculate $(\frac{\partial S}{\partial V})_T$, $(\frac{\partial S}{\partial T})_V$ and the second derivatives $(\frac{\partial^2 S}{\partial T \partial V})$.

[3] Ludwig Boltzmann (Vienna 1844–Duino (near Trieste) 1906, Austria (now Italy (suicide))) pioneered Statistical Mechanics with Gibbs.

Solution 5.4 The variation of S is

$$dS = \frac{\delta Q}{T} = \frac{dU + PdV}{T}.$$

From (5.38),

$$dU = u(T)dV + V\frac{\partial u}{\partial T}dT.$$

Also inserting (5.39), one finds

$$dS = \frac{\frac{4}{3}udV + V\frac{\partial u}{\partial T}dT}{T}. \tag{5.43}$$

From (5.43), $dS = (\frac{\partial S}{\partial V})dV + (\frac{\partial \mathrm{S}}{\partial T})dT$, where

$$\left(\frac{\partial S}{\partial V}\right) = \frac{4u}{3T},$$

while

$$\left(\frac{\partial S}{\partial T}\right) = \frac{V}{T}\frac{\partial u}{\partial T}.$$

According to the *second Principle, d S* must be exact; hence, the mixed derivatives must be equal:

$$\frac{\partial}{\partial T}\left(\frac{\partial S}{\partial V}\right) = \frac{\partial}{\partial V}\left(\frac{\partial S}{\partial T}\right).$$

Therefore,

$$\frac{4}{3T}\frac{\partial u}{\partial T} - \frac{4}{3}\frac{u}{T^2} = \frac{1}{T}\frac{\partial u}{\partial T},$$

that is,

$$\frac{\partial u}{\partial T} = 4\frac{u}{T}.$$

One can again derive the Stefan–Boltzmann law from this result.

Problem 5.5 The temperature of the solar photosphere is $T_S \sim 5800\,^{\circ}\mathrm{K}$ and the size of the solar disk seen from here is $\frac{1}{4}^{\circ}$. Evaluate the mean temperature of the Earth assuming that the Earth and the Sun are black bodies. What would the temperature of the Earth become if the distance from the sun were doubled?

Solution 5.5 The angle is $\frac{\pi}{4\times 180}$ rad, therefore, the fraction of the solid angle occupied by the Sun is $\Delta\Omega = \int d\phi d\cos(\theta) = \frac{2\pi(1-\cos(\alpha))}{4\pi} = 4.76\,10^{-6}$. By the Stefan–Boltzmann law, the power goes with T^4. Therefore the ratio of temperatures is

$\frac{T^E}{T_S} = 0.0467$. We obtain $T_E = 270.1\,^\circ$K, which is realistic (only slightly low, mainly because we neglected the greenhouse effect). Dividing α by 2 we should find 191.6 °K.

5.4 Classical Statistical Mechanics and Information

The results of many experiments performed on macroscopic bodies depend on the thermal state. The theoretical interpretation of these experiments in microscopic terms is the subject of Statistical Mechanics, which, unlike Thermodynamics, assumes prior knowledge of the interactions between the microscopic components of the system, i.e., the Hamiltonian $H(p, q)$. Moreover, one assigns certain parameters that determine the system from the macroscopic point of view, such as pressure, temperature, number of molecules for each species and so on. All kinds of measurements may be considered. Observables can be macroscopic, like the pressure of a gas as a function of the temperature or its specific heat, but all spectroscopies, including those designed to determine the microscopic properties of solids, are influenced by temperature effects and must be interpreted through statistical mechanics.

For example, Stern and Zartmann, in a famous experiment in 1920, measured the speed distribution of the molecules of a gas. The gas is contained in a tube; a small hole in the bottom of the tube is opened for a very short time and the molecules come out with the speed they had within the gas. The molecular beam so obtained impinges on a quickly spinning wheel, and the molecules arrive in different points depending on the time of flight. The velocity distribution, that one can convert into an energy distribution, is an asymmetric bell-shaped distribution whose maximum moves to higher energies by increasing temperature. Thus, Zartmann verified the law established theoretically by Maxwell[4] (in obvious notation)

$$dN = N\frac{2}{\sqrt{\pi}}\frac{1}{(K_B T)^{3/2}} e^{-\frac{\epsilon}{K_B T}}\sqrt{\epsilon}\, d\epsilon. \tag{5.44}$$

We shall derive this law in Sect. 5.8.2.

Most of the $N_A \sim 6 \times 10^{23}$ molecules in a mole have a kinetic energy of order $K_B T$ but a few are much slower or much faster than that. Every macroscopic measurement of some magnitude $f(p, q)$ which depends on the same arguments as $H(p, q)$ returns its average $\bar{f}$; the microscopic state cannot be measured and is therefore irrelevant. One cannot evaluate the evolution starting with an integral of the canonical equations. The proper use of the Hamiltonian was found by Gibbs.[5]

[4] Note that $\int_0^\infty \sqrt{x} e^{-x} dx = \frac{\sqrt{\pi}}{2}$, and so $\int dN = N$.

[5] The great theoretician Josiah Willard Gibbs (New Haven, Connecticut, U.S.A., 1839–Yale 1903) was the first professor of Mathematical Physics in the U.S.A.

5.5 Gibbs Averages

Suppose our macroscopic system is a mole of some substance that we wish to treat as a mechanical system; a point γ of the *phase space* or Γ *space* has $2s$ coordinates (s are space coordinates and s the conjugated momenta). So, the coordinates

$$q_1, q_2, \ldots, q_s, p_1, p_2, \ldots, p_s$$

of γ are just the arguments of the Hamiltonian. This huge list of variables specifies the microscopic state (or simply *microstate*) of the system. The evolution of the system is described by the chaotic motion of $\gamma(t)$.

The outcome of the macroscopic measurement of some magnitude $f(\gamma)$ does not depend on any details of $\gamma(t)$, and many microstates would give the same outcome. There are only small fluctuations, which scarcely perturb a regular behavior. If the measurement is done in equilibrium, and the system is perturbed in any way, the equilibrium can be reached again after a characteristic waiting time. If the measurement is then repeated, the measurement yields the equilibrium results again, without any memory of the twists of $\gamma(t)$ in Γ space. Therefore, obtaining $\gamma(t)$ from $\gamma(0)$ is not only impossible, but also irrelevant. What one measures is better represented by a time average, in obvious notation,

$$\overline{f(p,q)} = \frac{1}{T}\int_0^T f(p,q)dpdq, \ T \to \infty,$$

and the above discussion strongly suggests that $\bar{f}$ must be independent of the initial condition $\gamma(0)$. The time average would wash out all the information gained by the integration of the equations of motion and any distinction between the many microstates that would give the same results.

We must do without detailed initial conditions, since the system evolves spontaneously toward equilibrium[6] within some characteristic times.

The fundamental approach of Statistical Mechanics is credited to J. W. Gibbs, and the basic principle is very simple:

> *The fundamental assumption of Statistical Mechanics is: if a system in equilibrium has energy E, all the microscopic states of energy E are equally likely.*

Consider a system with Hamiltonian $H(p,q)$ in a state specified by all the relevant *macroscopic* parameters, like pressure, volume and number of particles. What can we say about its *microscopic* state? Gibbs considered the probability dP that it be in a given bunch $d\Gamma$ of microstates; more precisely,

$$dP = \rho(p,q)d\Gamma, \tag{5.45}$$

[6] If the system is the Universe, it is quite clear that it is not in thermal equilibrium; however, who knows if it evolves toward an equilibrium of some kind?

where $d\Gamma = \prod_i^s dp_i dq_i$ is the volume element of *phase space* Γ of the system. ρ is the *distribution function* . It must be normalized, that is,

$$\int_{\Gamma\text{space}} \rho(p,q)d\Gamma = 1.$$

Thus, the average of any magnitude f is

$$\bar{f} = \int f(p,q)\rho(p,q)d\Gamma. \tag{5.46}$$

A priori all the microstates are equally likely; therefore, all the portions of phase space compatible with the macroscopic parameters must be treated in the same way. $H(p,q)$ determines $\rho(p,q)$ and thus the probability of any given element $d\Gamma$. Put another way, one considers an *ensemble* of systems with the same H and with given macroscopic parameters; they are distributed in Γ space according to $\rho(p,q)$; we perform a statistical average on this *ensemble*.

Is the *ensemble* average really equal to the time average of the corresponding quantity that one would compute from the microscopic evolution? This is a very good question, which is best known as the *ergodic problem*. Physicists use this as a sort of postulate, but from the formal mathematical viewpoint, it is a difficult issue, which has been settled by Fermi and others, but only under certain assumptions. However one fact is true: based on this assumption, Statistical Mechanics works.

Liouville Theorems

According to Eq. (5.46), the distribution function $\rho(p,q)$ must be the probability if finding the system in $d\Gamma$. An apparent difficulty arises because during the evolution of the system the variables p and q evolve according to the Hamilton equations. Does the probability change with time? A Liouville theorem states that there is no problem, since

$$\frac{d\rho}{dt} = 0. \tag{5.47}$$

A second Liouville theorem states that given a set of representative p,q points in Γ space at some instant of time, it evolves as an incompressible fluid. The hypersurface containing the points can change shape, but the occupied volume remains constant. The meaning of this statement is easily understood. We must check that the Gibbs theory does not break down with a change of Lagrangian coordinates (punctual transformation $(p,q) \to (P,Q)$). In other words,

$$\bar{f} = \int f(P,Q)\rho(P,Q)\prod_i^s dP_i dQ_i = \int f(p,q)\rho(p,q)\prod_i^s dp_i dq_i.$$

This requires that the Jacobian be 1, that is,

$$\frac{\partial(Q, P)}{\partial(q, p)} = 1 = \frac{\partial(q, p)}{\partial(Q, P)}.$$

In the case of a single degree of freedom, the result is obvious, since

$$\frac{\partial(Q, P)}{\partial(q, p)} = Det \begin{pmatrix} \frac{\partial Q}{\partial q} & \frac{\partial Q}{\partial p} \\ \frac{\partial P}{\partial q} & \frac{\partial P}{\partial p}, \end{pmatrix} = \{Q, P\} = 1$$

because the Poisson brackets are invariant, as shown in Chap. 2. Liouville has shown that this remains true in the presence of many degrees of freedom, and therefore

$$\int dqdp = \int dQdP,$$

that is, if you cut out a volume inside Γ, its extension is conserved during the evolution.

5.6 Microcanonical Ensemble

In order to accomplish the above task, we must specify an *ensemble*. The Microcanonical Ensemble is a collection of a huge number of isolated systems; they must be systems with the same number of particles and about the same energy E_0; the energy uncertainty δE must be so small that adding it to the system does not change the experiment appreciably, while it may well be large compared to a molecular energy.

The microstates (i.e., the microscopic details at the molecular level) are random. All the magnitudes, including the conserved energy E_0, can only be specified with macroscopic accuracy, and it is not necessary to specify δE further. Experimentally, it is very hard to work with isolated systems in equilibrium, but conceptually, this is a good starting point.

Recall that a point in Γ space represents a system (not a molecule). We need to find the distribution function $\rho(p, q)$ which is the probability density of systems in Γ space, proportional to the number of systems in the ensemble that have their position in an element $d\Gamma$ around the point (p, q). In a macroscopic description, $\delta E \to 0$ and ρ_{micro} must be proportional to $\delta(H(p, q) - E_0)$ and normalized to 1. Letting $\omega(E_0) = \int d\Gamma \delta(H(p, q) - E_0)$, we arrive at the microcanonical distribution function

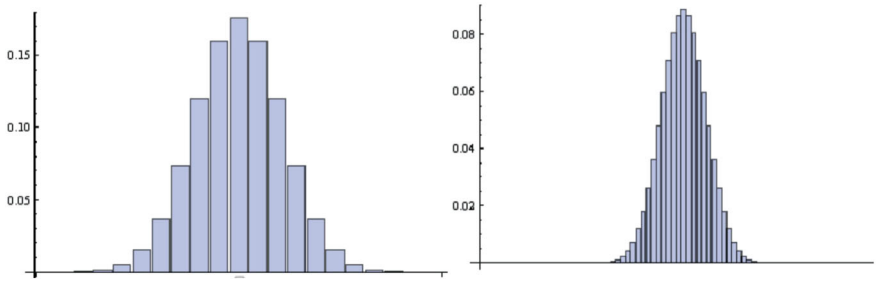


Fig. 5.2 Histograms of the probability distribution of N molecules in two equal boxes. Left: N=20, Right:N=80. Even with such ridiculous numbers, it is evident that the distribution becomes very sharp when N grows

$$\rho_{micro}(p,q) = \frac{\delta(H(p,q) - E_0)}{\omega(E_0)},\ \omega(E_0) = \int d\Gamma \delta(H(p,q) - E_0). \qquad (5.48)$$

Here, $\omega(E_0)$ is the measure of the constant energy surface. We must perform statistical averages over an *ensemble*, specifically a *microcanonical ensemble* of systems that are represented in Γ-space by points that are uniformly scattered in the gap $d\Omega$ between the hypersurfaces $H = E_0$ and $H = E_0 + \delta E$.

A simple example might help. Consider a perfect gas in equilibrium and suppose the volume of the container divided into two parts A and B with volumes V_A and V_B. In phase space, there is a set Γ of points that correspond to unbelievable situations in which all particles are in A. These points are treated on the same footing as the others. All ways of distributing the molecules are represented in the ensemble and the energy is the same for all of them. Imagine the molecules are numbered. The probability that number 1 is in A is $\frac{V_A}{V_B}$, and the same holds true for the others; the probability of having all in A is $(\frac{V_A}{V_B})^N$. Therefore, if $\frac{V_A}{V_B} = \frac{1}{2}$ and $N = 10^{23}$, the probability of having all in A is 10 to a power -3×10^{22}. This is practically 0 in Physics. The probability of having k molecules on one side and $N - k$ on the other side is $P(k,N) = 2^{-N}\begin{pmatrix} N \\ k \end{pmatrix}$, and is shown as a function of k in the next figure for $N = 20$ and $N = 80$. While fluctuations around the most probable value always exist, very large deviations are impossible. This is why the evolution of the system has a preferred direction, from unlikely to likely. This explains the arrow of time (Fig. 5.2).

Boltzmann generalized this concept and conceived a Γ space divided into small cells (the criterion of smallness will be made more precise below). Let $\delta\Gamma$ denote the volume of each cell,which corresponds to a microstate of the system, with all the q and p specified with a small uncertainty. The cells are finite, but so small that the uncertainty of q and p is much smaller than any characteristic length or momentum of the problem. Since $q \times p$ is an *action*, let the uncertainty of pq be a small action[7]

[7] Planck's quantum h had not yet been discovered.

h; then the volume of the cell is h^s, where s in the number of degrees of freedom of the system.

h is kept finite. In this way, each volume of Γ space corresponds to a very large number of cells. Let $\Omega_{tot}(E)$ be the volume of the region of Γ where the energy of the system is in the interval between 0 and E and the volume, pressure and other macroscopic parameters are also fixed; dividing the volume by per h^s, one obtains a very large number of microstates

$$W_{tot}(E) = \frac{\Omega_{tot}(E)}{h^s}.$$

This number is a measure of $\Omega_{tot}(E)$ for the given macroscopic conditions.

A small increment δE corresponds to a gap $\delta\Omega_{tot}(E_0)$ of phase space that contains some number $W(E_0)$ of cells.

W is the number of microstates that appear macroscopically as the same state of the system, with energy E.

Note that W is given by

$$W(E_0)h^s = \delta\Omega_{tot} = \left(\frac{\partial\Omega_{tot}}{\partial E}\right)\delta E = \omega(E_0)\delta E. \tag{5.49}$$

W is proportional to the measure of the constant-energy hypersurface

$$W = \frac{\partial W_{tot}}{\partial E}\delta E = \frac{\delta E}{h^s}\int \delta(H - E)d\Gamma. \tag{5.50}$$

The argument is very ingenious. It appears to *stress* the mathematics, but it is perfectly sound. Since δE is not infinitesimal, W is a large number. This appears to be a difficulty, since then W is defined only up to a multiplicative constant. If we can vary h, we must vary W accordingly. Nature actually defines a scale for h, which is Planck's constant, and that does not fit in the classical scheme.

Now we can compare two hypothetical states of a macroscopic system, both compatible with the macroscopic parameters. Compare, for example, the state with the gas that occupies all the available volume and the state in which the gas fills only one part of the partition. We have seen that the number W of ways in which the latter situation can occur is negligible. States with larger W are closer to equilibrium, which requires maximum W. The equilibrium states are those with maximum entropy S. This reasoning motivated Boltzmann to look for a functional link of the form

$$S = f(W).$$

S is extensive: the entropy of a system formed by two parts is the sum of the entropies $S = S_1 + S_2$, while by the laws of probability, $W = W_1 W_2$. Therefore, f must satisfy

$$f(xy) = f(x) + f(y).$$

That is enough to identify f with the logarithm: so Boltzmann proposed the fundamental equation

$$\boxed{S = K_B \log W = K_B \log \frac{\delta E}{h^s} \int \delta(H - E) d\Gamma} \tag{5.51}$$

that prompts a *statistical meaning* of entropy. This is the microcanonical definition of S. Remarkably, any change of h or δE, $\log W$ only changes an additive constant like in thermodynamics, but S is a state function. According to the enlightening equation (5.51), S is a measure of our ignorance of the microstate of the system. The system evolves by increasing the entropy; in a sense, it is shy and tries to hide the microstate by minimizing our knowledge of it. S increases with the degree of disorder. Only at absolute zero is the microstate unveiled because the system is compelled to reach the minimum energy state, and then $S = 0$ in agreement with the third principle of thermodynamics. However, the prudish system does not allow this, since one cannot reach absolute zero!

5.6.1 Entropy of the Perfect Gas

Consider a perfect gas of N atoms with mass m, and Hamiltonian

$$H = \sum_i^N \frac{p_i^2}{2m},$$

in a volume V. To obtain S by (5.51), we need the number of cells of volume h^{3N} in the set of Γ space where the energy is $\leq$E. The projection of this domain on the space of coordinates $q_1 \ldots q_{3N}$ is a hypercube with side $V^{1/3}$; its projection on the space with coordinates $p_1 \ldots p_{3N}$ is a hypersphere, and, therefore,

$$W_{tot}(E) = \frac{1}{h^{3N}} \int \prod_j^N d^3 p_j d^3 q_j \theta \left(2mE - \sum_i^N p_i^2 \right).$$

Integrating over q yields V^N, and we must work out

$$W_{tot}(E) = \left(\frac{V}{h^3} \right)^N \Omega_{3N}(\sqrt{2mE}),$$

where $\Omega_{3N}(R)$ is the volume of the hypersphere of radius R in 3 N dimensions; see (3.29). Consequently,

$$W_{tot}(E) \equiv \frac{\Omega_{tot}(E)}{h^{3N}} \frac{\pi^{\frac{3N}{2}}}{(\frac{3N}{2})!} \left(\frac{V}{h^3}\right)^N (2\,mE)^{\frac{3N}{2}} = \frac{1}{(\frac{3}{2}N)!} \left[\frac{V}{h^3}(2\pi mE)^{3/2}\right]^N. \tag{5.52}$$

So,

$$\log(W_{tot}(E)) = N \log\left[\frac{V}{h^3}(2\pi mE)^{3/2}\right] - \log\left[\left(\frac{3\,N}{2}\right)!\right].$$

Using Stirling's formula, one obtains

$$\log W_{tot} = N \log\left[V\left(\frac{4\pi m}{3\,h^2}\frac{E}{N}\right)^{\frac{3}{2}}\right] + \frac{3}{2}N. \tag{5.53}$$

According to (5.51), (5.49), (5.50) in order to calculate S, we must take the logarithm of $W = \frac{\partial W_{tot}}{\partial E}$. By (5.52),

$$W_{tot}(E) = \text{constant} \times E^{\frac{3N}{2}} \;\Rightarrow\; W(E)\frac{\partial W_{tot}}{\partial E} = \frac{3\,N}{2E} W_{tot}.$$

So,

$$\log W(E) = \log W_{tot} + \log N + \text{constant}.$$

Recalling (5.49) and (5.50), this result has a geometrical content. In terms of the volume $\Omega(E)$ of the hypersphere, it gives the measure $\omega(E)$ of its hypersurface. By Eq. (5.53), the logarithm of $W_{tot}(E)$ for large N is dominated by the terms in $N \ln N$ plus terms linear in N; the factor that allows us to transform $\Omega(E)$ into $\omega(E)$ is a logarithm, which diverges for $N \to \infty$, but much more slowly than the dominant terms. For this reason, the logarithm is negligible in the thermodynamic limit. The constant δE is even more negligible. To sum up, $\log \omega(E)$ and $\log \Omega_{tot}$ are equal in the thermodynamic limit $\mathrm{N} \to \infty$. Therefore, we take $\Omega(E)$ instead of $\omega(E)\delta E$. Thus,

$$\frac{S}{K_B} = \log W_{tot}(E) = N \log\left[V\left(\frac{4\pi m}{3\,h^2}\frac{E}{N}\right)^{\frac{3}{2}}\right] + \frac{3}{2}N. \tag{5.54}$$

To test the validity of the present statistical approach, we can compare it with the result (5.34), from thermodynamics. Does it agree? No!!

The Paradox

Gibbs noted that this *honest* calculation (done by applying the known rules) leads to a paradox. The entropy is an *extensive* quantity, that is, it is proportional to N. The expression (5.54) for S is surely wrong since it is not extensive. So the rules must be changed, but how? To fix this severe problem, Gibbs proposed a new and very far-reaching physical idea. He suggested that the phase space for a system of N identical molecules is not $\Omega(E)$, but $\frac{\Omega(E)}{N!}$ (in the usual action units). Now, $N! = N(N-1)(n-2)\ldots 3.2.1$ is the number of permutations on N objects. All the microstates that differ by permutations of the particles are identified by Gibbs,

that is, they are taken as the same configuration. Classically the exchange of two like atoms should lead to a different microstate, since it is possible (in principle, not in reality) to observe them continuously as they move and exchange place; at least, it is possible to mark them somehow, in the same way as billiard balls are marked by chalk without a visible effect on their trajectories. According to Gibbs, atoms cannot be treated like tiny billiard balls. Then, $S \to S - \log N! \sim S - (N \log N - N) = S - N \log(\frac{N}{e}) = S + N \log(\frac{e}{N})$, and we get the additive result

$$\frac{S}{K_B} = \log \Omega(E) = N \log \left[\frac{eV}{N} \left(\frac{4\pi m}{3h^2} \frac{E}{N} \right)^{\frac{3}{2}} \right] + \frac{3}{2} N \tag{5.55}$$

in agreement with Thermodynamics. Gibb's intuition anticipated the quantum indistinguishability of identical particles!

5.7 Canonical Ensemble

The basic principle of Statistical Mechanics affirms the equal probability of the microstates; this idea is realized most directly in the microcanonical ensemble. However, in order to be able to assign a sharp energy to the microstates, we must assume that the system does not exchange heat or work with anything. This is a problem, since experiments are never done on isolated systems. Keeping a system effectively isolated is very difficult, if not impossible, and interesting problems usually deal with samples that are in equilibrium with a thermal bath at some temperature.[8] In the Canonical Ensemble the system defined by the Hamiltonian $H(p, q)$ is in equilibrium at temperature T. The system can be large or small, even a single molecule. Some an interaction with a thermal bath is needed to fix T, but the coupling must be so small that we can continue to think in terms of $H(p, q)$. These requirements appear to be conflicting, but they are not really, since the coupling to the external bath can be very small. In practice it will take a very long time to establish equilibrium by a tiny interaction, but this is equilibrium Statistical Mechanics and there are no time constraints.

5.7.1 Canonical Temperature

When a body is in equilibrium with a thermal bath, its internal energy fluctuates. There are exchanges of thermal energy $dE = \delta Q$. Even the entropy fluctuates. In

[8] The temperature is defined in terms of a Carnot cycle, therefore the heat sources must be so large that the Carnot cycle does not change their thermal state. A system at a fixed temperature exchanges energy with a heat bath, therefore its energy content fluctuates; in the microcanonical ensemble the energy is fixed, and therefore the temperature fluctuates. However, for a macroscopic sample, such fluctuations are relatively unimportant. Therefore it is expected that both schemes should lead to the same results for large systems in most cases.

thermodynamics, $dS = \frac{\delta Q}{T}$. Up to now, we have a statistical definition of S, but not yet of temperature T. In the microcanonical ensemble, a microstate has fixed E, but T is not sharp. A canonical definition of temperature is inherent in the above remarks. Let us introduce the inverse canonical temperature

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right); \tag{5.56}$$

this links energy fluctuations to entropy fluctuations in agreement with equation (5.56). Now let us look at the equilibrium condition. Suppose an isolated system has two weakly interacting subsystems:

$$H \sim H_1 + H_2, \quad d\Gamma = d\Gamma_1 d\Gamma_2.$$

They cannot have a sharp energy content, since there are energy fluctuations dE_1, dE_2, but, since the mutual interaction is negligible, $dE_1 + dE_2 = 0$. The equilibrium condition is that S must be maximum. Therefore,

$$dS = \left(\frac{\partial S_1}{\partial E_1}\right) dE_1 + \left(\frac{\partial S_2}{\partial E_2}\right) dE_2 = 0.$$

Thus, the equilibrium is given by

$$\left(\frac{\partial S_1}{\partial E_1}\right) = \left(\frac{\partial S_2}{\partial E_2}\right).$$

This is in line with the definition (5.56) of the temperature.

5.8 Canonical Distribution

In the Canonical Ensemble, the system under consideration $\mathcal{S}$ and the thermal bath $\mathcal{B}$ make up an isolated[9] system $\mathcal{C}$ with Hamiltonian[10]

$$H_{\mathcal{C}} \sim H_{\mathcal{S}} + H_{\mathcal{B}}. \tag{5.57}$$

Therefore, the energy of the system that we wish to study is not fixed, so we shall replace the distribution function ρ_{micro} with a different one $\rho(p, q)$.

[9] It is evident that possible interactions of other systems with the thermal bath are not relevant if the bath is very large.

[10] The careful reader could object that an essential term is missing: the interaction of our system with the bath. However, such interaction term would modify the system, and on the other hand, in the long run, an equilibrium is granted even by such a small interaction that its effect on the dynamics of the system can be neglected. The argument is subtle.

Now, $\mathcal{C}$ has a fixed energy $E_{\mathcal{C}}$ and cannot have a sharply defined temperature, but we may deal with it in the microcanonical scheme. The element of the phase space is obviously the product $d\Gamma_{\mathcal{C}} = d\Gamma_{\mathcal{S}} d\Gamma_{\mathcal{B}}$ and the probability of finding $\mathcal{C}$ in $d\Gamma_{\mathcal{C}}$ is

$$dP_{\mathcal{C}} = \frac{\delta(H_{\mathcal{B}} + H_{\mathcal{S}} - E_{\mathcal{C}})}{\int \delta(H_{\mathcal{B}} + H_{\mathcal{S}} - E_{\mathcal{C}}) d\Gamma_{\mathcal{C}}} d\Gamma_{\mathcal{B}} d\Gamma_{\mathcal{S}}. \tag{5.58}$$

The large bath has an enormous number of degrees of freedom and is essentially arbitrary. Therefore, this microcanonical calculation is impossible to do, and very uninteresting. The clever idea is not to compute (5.58), but to integrate out the bath. We note that the denominator produces some multiplicative constant which depends on the uninteresting energy $E_{\mathcal{C}}$; since at the end we must impose a normalization condition, this constant will be thrown away, and so we shall forget about it. Therefore, the probability of finding S in its volume element is

$$dP_{\mathcal{S}} = C d\Gamma_{\mathcal{S}} \int \delta(H_{\mathcal{B}} + H_{\mathcal{S}} - E_{\mathcal{C}}) d\Gamma_{\mathcal{B}}, \tag{5.59}$$

with C to be determined by normalization.

Essentially the same integral appears in the microcanonical entropy (5.51) of the isolated $\mathcal{B}$:

$$S_{\mathcal{B}}(E) = K_B \log(W_{\mathcal{B}}(E)) = K_B \log\left[\delta E \int \delta(H_{\mathcal{B}} - E) \frac{d\Gamma_{\mathcal{B}}}{h^{s(\mathcal{B})}}\right], \tag{5.60}$$

where $s(\mathcal{B})$ is the number of degrees of freedom of the bath and $E = E_{\mathcal{C}} - H_{\mathcal{S}}$. Here, δE adds a constant to the entropy and can be ignored. Exponentiating,

$$e^{\frac{S_B(E)}{K_B}} = \int \delta(H_B - E) \frac{d\Gamma_{\mathcal{B}}}{h^{s(\mathcal{B})}}.$$

In Eq. (5.58), however, $E = E_{\mathcal{C}} - H_{\mathcal{S}}$. Therefore, we need

$$\int \delta(H_{\mathcal{B}} - E) \frac{d\Gamma_{\mathcal{B}}}{h^{s(\mathcal{B})}} = e^{\frac{S_{\mathcal{B}}(E_{\mathcal{C}} - H_{\mathcal{S}})}{K_{\mathcal{B}}}}. \tag{5.61}$$

Since the bath is very large, we can expand

$$S_{\mathcal{B}}(E_{\mathcal{C}} - H_{\mathcal{S}}) \sim S_{\mathcal{B}}(E_{\mathcal{C}}) - \frac{\partial S_{\mathcal{B}}}{\partial E} H_{\mathcal{S}}$$

and in view of (5.56),

$$S_{\mathcal{B}}(E_{\mathcal{C}} - H_{\mathcal{S}}) \sim S_{\mathcal{B}}(E_{\mathcal{C}}) - \frac{H_{\mathcal{S}}}{T}.$$

Therefore,

$$dP_{\mathcal{S}} = Ce^{-\frac{H_{\mathcal{S}}(p,q)}{K_B T}} d\Gamma_{\mathcal{S}} \equiv \rho d\Gamma_{\mathcal{S}}.$$

The *canonical distribution* for $\mathcal{S}$ is the dimensionless $\rho(p,q) = Ce^{-\frac{H_{\mathcal{S}}(p,q)}{K_B T}}$. But from now in, we can simplify the notation dropping S throughout, writing $d\Gamma$ for $d\Gamma_{\mathcal{S}}$, since $\mathcal{B}$ is eliminated. We also use the standard notation

$$\beta = \frac{1}{K_B T}. \tag{5.62}$$

So,

$$\boxed{\rho = \frac{e^{-\beta H(p,q)}}{\int e^{-\beta H(p,q)} d\Gamma.}} \tag{5.63}$$

The denominator is the *partition function* (which ensures that the integral of ρ over Γ is 1),

$$Z = \int e^{-\beta H(p,q)} d\Gamma, \tag{5.64}$$

which is a central quantity, as we shall see. The canonical average of any magnitude $A(p,q)$ is

$$\overline{A} = \frac{1}{Z} \int e^{-\beta H(p,q)} A(p,q) d\Gamma. \tag{5.65}$$

If the system is formed by two independent parts, $H_{\mathcal{S}} = H_1 + H_2$, and thus $\rho = \rho_1 \rho_2$, the parts 1 and 2 have their own canonical distribution.

The canonical definition of the Free Energy is

$$F = -K_B T log(Z), \tag{5.66}$$

while the canonical entropy, valid for any system, reads as

$$\boxed{\frac{S_m}{K_B} = -\sum_i p_i \log(p_i),} \tag{5.67}$$

where p_i is the probability of finding the system in the microstate i. Next we examine how the canonical distribution works in the simple case of a perfect gas,

5.8.1 Hint to Shannon Entropy and Information Theory

In 1948, Claude Shannon found a new important use of the expression (5.67) of the canonical entropy, and started the theory of information. Suppose the quantities

p_i represent the information content of the $i-$th element of a message and $\log(p_i)$ represents the probability of occurrence of this element on a message. In this context, the basis of the logarithm is taken to be 2. Then,

$$\frac{S_m}{K_B} = -\sum_i p_i \log(p_i)$$

with i running over the elements of the messages (like sentences) measures the information content.

The starting point is the definition of the entropy (that is, information content) associated with a single event, $I = -ln(p)$, where p is the probability that the event happens. Clearly if the event is highly probable, p is close to 1 and I is very small; so I is a measure of the amount of surprise, or information, connected with the news of the event. Starting with these premises, Shannon developed a theory of communication, which is important to quantify the quality of communication channels, and for many other applications including measuring the efficiency of an alphabet, data compression. This argument is not directly relevant to a book about theoretical physics, but is surely one of the important developments of Statistical Mechanics and is worth briefly mentioning here.

5.8.2 Maxwell Distribution of Molecular Energies

In the case of a perfect gas in a volume V, the gas is the isolated system $\mathcal{C}$ of the previous section, while an atom is $\mathcal{S}$; then, $H = \frac{p^2}{2m}$, $d\Gamma = d^3x d^3p$, and so,

$$dP = C \exp\left[-\frac{p^2}{2mK_BT}\right] d^3x d^3p. \tag{5.68}$$

Since $\int d^3x$ is the volume V the normalization condition is

$$CV \int \exp\left[-\frac{p^2}{2mK_BT}\right] d^3p = 1.$$

Inserting $\int_{-\infty}^{\infty} dx e^{-\alpha x^2} \sqrt{\frac{\pi}{\alpha}}$,

$$CV = \frac{1}{(2\pi m K_B T)^{3/2}} \Leftrightarrow Z = V(2\pi m K_B T)^{3/2}.$$

The velocity distribution is obtained by $(p_i \to mv_i)$ and reads as

$$dP(\boldsymbol{v}) = \left(\frac{m}{2\pi K_B T}\right)^{2/3} \exp\left(\frac{1}{2}\frac{mv^2}{K_B T}\right) dv_x dv_y dv_z.$$

There is spherical symmetry, so one sets $d^3v = v^2 dv \sin\theta d\theta d\phi$ and integrates over the angles. The result is

$$dP(v) = 4\pi \left(\frac{m}{2\pi K_B T}\right)^{2/3} \exp\left(\frac{1}{2}\frac{mv^2}{K_B T}\right) v^2 dv. \tag{5.69}$$

The Maxwell distribution (5.44) can be obtained by inserting $\frac{1}{2}mv^2 = \epsilon$, with $dv = \sqrt{\frac{1}{2m\epsilon}} d\epsilon$, and the number of molecules N. Indeed, we find

$$dN = N \frac{2}{\sqrt{\pi}} \frac{1}{(K_B T)^{3/2}} e^{-\frac{\epsilon}{K_B T}} \sqrt{\epsilon} d\epsilon.$$

5.8.3 *Perfect Gas in the Canonical Ensemble and Boltzmann Statistics*

Above, we derived the Maxwell distribution from the canonical distribution. In this section, we check that it arises directly from Gibb's principle that all the microstates must have the same probability. The perfect gas consists of $N >> 1$ independent material points with total energy E in equilibrium. The gas is the thermal bath $\mathcal{B}$ and $\mathcal{S}$ is just the molecule. The whole system $\mathcal{C} = \mathcal{B} + \mathcal{S}$ is treated as in the microcanonical ensemble, with fixed energy and particle number. All the microstates of energy E are taken to be equally likely. We work out the distribution $n(\epsilon)$ of the single molecule energy ϵ that grants the maximum of the number W of microstates of $\mathcal{C}$. In short, the equilibrium distribution is the one that can be realized in more ways.

Since the molecules are independent, instead of the Γ space we can work in the single molecule μ space, with coordinates x, y, z, p_x, p_y, p_z. We divide μ space in cells of volume $\Delta x \Delta y \Delta z \Delta p_x \Delta p_y \Delta p_z$. In this way we can discretize. The energy axis ϵ of a molecule will be sliced in small intervals of microscopic width $\Delta\epsilon$. The exact method of slicing is irrelevant as long as the cells are macroscopically small and the energy $\Delta\epsilon$ is sufficiently small. However. the cells must still be so large that each cell contains several molecules.

Each way to assign the individual molecules to the cells defines a microstate. We shall treat the molecules as distinguishable as if each molecule could be labeled by putting a number over it. This looks like an obvious assumption in classical physics that we shall drop in Quantum Mechanics. Therefore, classically the exchange of the μ space coordinates (p, q) of two molecules takes to a distinct microstate. The equilibrium canonical distribution

$$\overline{n}_i = N\rho(\epsilon_i) \tag{5.70}$$

must be normalized according to

$$\sum_i \rho(\epsilon_i) = 1. \tag{5.71}$$

The probability of any distribution $n_1, n_2, n_3 \ldots$, with n_i molecules in ith cell, is proportional to the number of ways it can be realized, compatibly with the restrictions

$$\sum n_i = N$$

and

$$\sum \epsilon_i n_i = E.$$

The number of ways[11] to choose n_1 particles to put in cell 1 is

$$\binom{N}{n_1};$$

when this is done, one is left with $N - n_1$ molecules from which one can choose n_2 for the second cell. A distribution $n_1, n_2, n_3, \ldots$ is realized in

$$W = \binom{N}{n_1}\binom{N-n_1}{n_2}\binom{N-n_1-n_2}{n_3}\ldots \tag{5.72}$$

different ways. Here a colossal simplification occurs, since

$$\begin{aligned}
&\binom{N}{n_1}\binom{N-n_1}{n_2}\binom{N-n_1-n_2}{n_3} \\
&= \frac{N!}{n_1!(N-n_1)!}\frac{(N-n_1)!}{n_2!(N-n_1-n_2)!}\frac{(N-n_1-n_2)!}{n_3!(N-n_1-n_2-n_3)!}\ldots\ldots \\
&= \frac{N!}{n_1!n_2!n_3!\ldots},
\end{aligned}$$

and so,

$$W = \frac{N!}{\prod_r n_r!}. \tag{5.73}$$

Since the logarithm is everywhere an increasing function of its argument, we maximize $\log(W)$; by the Stirling formula

[11] $\binom{N}{n} = \frac{N!}{n!(N-n)!}$ is the number of different choices of n objects from N, regardless of the order of the selected objects.

$$\ln W \approx N \ln N - \underbrace{N} - \sum n_i \ln n_i + \underbrace{\sum n_i} = N \ln N - \sum n_i \ln n_i. \quad (5.74)$$

This is the entropy of the gas

$$\frac{S}{K_B} = \ln W \approx N \ln N - \underbrace{N} - \sum n_i \ln n_i + \underbrace{\sum n_i} = N \ln N - \sum n_i \ln n_i \quad (5.75)$$

in terms of the (yet unknown) cell populations. According to the Lagrange method, we impose

$$\frac{\partial}{\partial n_r}\left[-\sum n_i \ln n_i - \gamma \sum n_i - \beta \sum \epsilon_i n_i\right] = 0, \quad (5.76)$$

where β, γ are multipliers. We get

$$-\ln n_r - 1 - \gamma - \beta\epsilon_r = 0,$$

that is,

$$\boxed{\overline{n}_r = e^{-\alpha} e^{-\beta\epsilon_r}} \quad (5.77)$$

with $\alpha = 1 + \gamma$, $\beta = \frac{1}{K_B T}$. Now we can grasp better the meaning of the Boltzmann distribution (5.77). The parameter α is fixed by (5.71). To obtain the entropy, we cast (5.70), (5.71) into (5.75):

$$\frac{S}{K_B} = N \log N - N \sum_r \rho(\epsilon_r) \log(N\rho(\epsilon_r)) - N \sum_r \rho(\epsilon_r) \log(\rho(\epsilon_r)).$$

The molecular distribution function reads as

$$\rho(\epsilon_r) = \frac{e^{-\beta\epsilon_r}}{Z}.$$

The canonical entropy per molecule of the gas is

$$\frac{S_m}{K_B} = -\sum_r^{cells} \rho(\epsilon_r) \log(\rho(\epsilon_r)). \quad (5.78)$$

This expression for the canonical entropy is in Eq. (5.67)

$$\boxed{\frac{S_m}{K_B} = -\sum_i p_i \log(p_i)), \quad (5.79)}$$

where p_i is the probability of finding the system in the microstate i.

The canonical partition function reads as

$$Z = \sum_r e^{-\beta\epsilon_r}, \tag{5.80}$$

and the canonical mean of any quantity A is

$$\overline{A} = \frac{1}{Z}\sum_r^{cells} e^{-\beta\epsilon_r} A(r). \tag{5.81}$$

Problem 5.6 A *mini-gas* of 10 molecules has 4 states (or cells) available. The cell energies are $0, \epsilon, 2.\epsilon, 3.\epsilon$. The total energy is $E = 14\epsilon$. Let n_k denote the number of molecules in the kth cell of energy $k\epsilon$. Find the most probable value of n_0.

Solution 5.6 There are 18 possibilities. Below I report the cell occupations:
((0, 6, 4, 0), (0, 7, 2, 1), (0, 8, 0, 2), (1, 4, 5, 0),
(1, 5, 3, 1), (1, 6, 1, 2), (2, 2, 6, 0), (2, 3, 4, 1),
(2, 4, 2, 2), (2, 5, 0, 3), (3, 0, 7, 0), (3, 1, 5, 1),
(3, 2, 3, 2), (3, 3, 1, 3), (4, 0, 4, 2), (4, 1, 2, 3),
(4, 2, 0, 4), (5, 0, 1, 4)).
The maximum probability occurs for $n_0 = 2$ and for $n_0 = 3$.

The thermodynamic internal energy per molecule U is identified with the average energy

$$U = \overline{H_S},$$

and is obtained from Z:

$$\overline{H} = -\frac{1}{Z}\frac{\partial Z}{\partial\beta} = -\frac{\partial \ln(Z)}{\partial\beta}. \tag{5.82}$$

The perfect gas example suggests to define the entropy per molecule

$$\frac{S_m}{K_B} = -\sum_r^{cells} \rho(\epsilon_r)\log(\rho(\epsilon_r)).$$

The canonical distribution (5.63) yields

$$\frac{S_m}{K_B} = -\sum_r \rho(\epsilon_r)[-\beta\epsilon_r + \log(Z)] = \frac{U}{NK_BT} + \log(Z). \tag{5.83}$$

From (5.8) $F = U - TS$, the molecular free energy reads

$$F = -K_BT\ln(Z).$$

Moreover, combining (5.83) with (5.82), we find

$$S_m = -\frac{1}{T}\frac{\partial}{\partial\beta}\log(Z) + K_B \log(Z),$$

and since

$$\frac{\partial}{\partial\beta} = -\frac{1}{K_B\beta^2}\frac{\partial}{\partial T},$$

the molecular entropy is

$$S_m = K_B T \frac{\partial}{\partial T}\log(Z) + K_B \log(Z)$$

or, equivalently,

$$S_m = \frac{\partial}{\partial T}[K_B T \ln(Z)] = -\left(\frac{\partial F(V,T)}{\partial T}\right)_V, \tag{5.84}$$

in accord with (5.9); recalling (5.10), it is easy to obtain the pressure

$$P = -\left(\frac{\partial F}{\partial V}\right)_T. \tag{5.85}$$

So, from the partition function all the thermodynamic quantities are readily obtained.

Problem 5.7 For the monoatomic perfect gas, calculate the partition function Z, the internal energy U, entropy S and pressure P.

Solution 5.7

$$Z = \frac{1}{h^3}\int d^3p d^3x \; exp\left[-\frac{p^2}{2mK_BT}\right] = \frac{1}{h^3}V(2\pi m K_B T)^{\frac{3}{2}}.$$

Hence,

$$U = -N\frac{\partial}{\partial\beta}\log Z = \frac{3}{2}NK_BT.$$

From (5.83),

$$\frac{S_m}{K_B} = \log\left(\frac{V(2\pi m K_b T)^{\frac{3}{2}}}{h^3}\right),$$

up to an additive constant

$$F = -K_B T \log(\frac{V(2\pi m K_B T)^{\frac{3}{2}}}{h^3}).$$

Finally, from (5.85),

$$P = \frac{K_B T}{V}.$$

5.8.4 Theorem of Equipartition

Consider the harmonic oscillator

$$H(p,q) = \frac{p^2}{2m} + V(q),\ \ V(q) = \frac{1}{2} m\omega^2 q^2.$$

Since

$$V(q) = \frac{1}{2} q \frac{\partial H}{\partial q},$$

the thermal average of V at temperature T is given by

$$\overline{V} = \frac{1}{2Z} \int dqdpq \frac{\partial H}{\partial q} e^{-\beta H} = \frac{1}{2Z} \int dqdp \frac{q}{(-\beta)} \frac{\partial}{\partial q} e^{-\beta H};$$

and by an integration by parts one finds

$$\overline{V} = \frac{-K_B T}{2Z} \int dqdp \left\{ \frac{\partial}{\partial q} (q e^{-\beta H}) - e^{-\beta H} \right\}.$$

This can be simplified, since $H \to \infty$ for $q \to \pm\infty$ and the first contribution vanishes; we are left with

$$\overline{V} = \frac{K_B T}{2}.$$

Similarly, one can obtain

$$\overline{T} = \frac{K_B T}{2}.$$

The traditional statement is that for any system of oscillators, there is a contribution $\frac{K_B T}{2}$ for each degree of freedom, i.e., from each q and each p. Systems of oscillators include the vibrations of molecules and solids and the normal modes of the electromagnetic field in a cavity. The failure of this theorem for the electromagnetic field is called ultraviolet catastrophe, and it led Planck to introduce the quanta (see below). But here, we go ahead with the classical theory.

Example: Monoatomic Perfect Gas

Since

$$H = \sum_i^N \frac{p_{xi}^2 + p_{yi}^2 + p_{zi}^2}{2m},$$

we rederive the internal energy is

$$U = \frac{3}{2} N K_B T$$

and the specific heat at constant volume $C_V = (\frac{\partial U}{\partial T})_V = \frac{3}{2} N K_B$. For one mole of gas, $N \sim 6.022 \times 10^{23}$ is the Avogadro number and $C_V = \frac{3}{2} R$, where $R = N K_B$ is the gas constant. These well-known results are all wrong at low temperatures: in particular, the specific heats vanish for $T \to 0$ for the third principle of Thermodynamics.

Example: Biatomic Gas

For biatomic molecules, if one considers only the translational degrees of freedom, the result is $C_V = \frac{3}{2} N K_B$ as above. But one should add to the kinetic energy the rotational Hamiltonian, which can be roughly represented by a rigid rotor

$$H_{\text{rot}} = \frac{1}{2I} \left(P_\theta^2 + \frac{1}{\sin^2 \theta} P_\phi^2 \right).$$

Due to the quadratic form, H_{rot} adds another R per mole. This is not all. The rotator is not rigid, therefore one should add a (roughly) harmonic potential and another R. Experimentally, e.g., for H_2, well above the boiling point of liquid Hydrogen, $C_V = \frac{3}{2} N K_B$ and the only degrees of freedom that contribute are translational, while the others are frozen. They start to contribute at higher temperatures. The explanation is possible only in the quantum theory.

Dulong and Petit Law

Einstein modeled the vibrations of an elemental solid as a harmonic oscillator associated with each of the N atoms. Thus, H is the sum of N oscillator terms and classically $C_V = 3R$. This agrees with a law by Dulong and Petit, which is about right around room temperature. However, at low temperatures, C_V drops, and this can be understood by Quantum Mechanics. Actually, Einstein had proposed a qualitatively correct explanation in terms of vibration quanta already in 1907 (see Chap. 25).

5.9 Ultraviolet Catastrophe and the Planck Formula

The spectrum, that is the dependence of energy density on frequency $u(\nu)$, was measured through its link (5.37) with the emissive power. At each temperature, the data show a bell-shaped $u(\nu)$ with a characteristic asymmetry that grows from 0 to a maximum and then decreases to zero more slowly at high frequencies. The maximum shifts to a higher frequency with increasing the temperature T and the maximum frequency is proportional to T (Wien's law). Rayleigh and Jeans, in 1900, treated the problem as follows. In a large cubic cavity of volume V, the number of right-polarized modes of the electromagnetic field with wave vector in d^3k is

$$d\mathcal{N} = \frac{V d^3k}{(2\pi)^3};$$

this must be multiplied by 2 to include left polarization. Integrating over the angles of $\overrightarrow{k}$ s by $\int_\Omega d^3k = 4\pi k^2 dk$ and setting $|k| = \frac{\omega}{c}$, one finds

$$\int_\Omega d^3k = \frac{4\pi\omega^2 d\omega}{c^3}; \tag{5.86}$$

finally,

$$d\mathcal{N} = 2\frac{Vk^2 dk}{\pi^2} = V\frac{\omega^2 d\omega}{\pi^2 c^3}. \tag{5.87}$$

Each mode is a harmonic oscillator and has energy $K_B T$ according to the equipartition theorem. The energy per unit volume is

$$u(\omega)d\omega = \frac{K_B T\omega^2}{\pi^2 c^3} d\omega. \tag{5.88}$$

This result is in excellent agreement with experiments in the region of low frequencies (well below the maximum), and this represents a great success of the above theory. But, when ω increases, things go so disastrously that each physicist refers to this problem as *the ultraviolet catastrophe* (Fig. 5.3).

The calculation of the number of modes is simple geometry and cannot be wrong. It is the equipartition theorem that would cause the absurd result $\int_0^\infty u(\omega)d\omega = \infty$, and each body should emit infinite energy. This can be traced back to the fact that the Boltzmann statistics does not apply to the states of the field and those of energy above $K_B T$ are effectively quenched. Again, the problem is deep and the solution requires Quantum Mechanics, which explains in detail the law of Wien and fixes the constant of the Stefan–Boltzmann law.

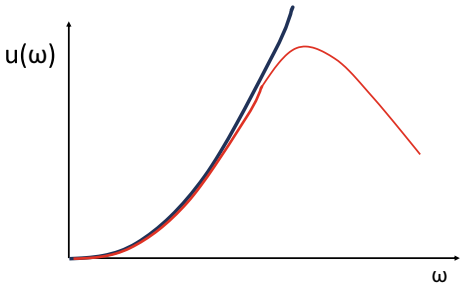


Fig. 5.3 Qualitative illustration of the ultraviolet catastrophe. The Raleigh–Jeans calculation gives a parabolic trend (black) which agrees well with the data at low frequency, but is disastrously wrong when the experimental data reach a maximum and then decrease

Planck in 1900 understood that the theory was fundamentally right, but the problem lay in how the various normal modes entered the theory, and specifically in how they were weighted. He saw that the problem was that the classical formula gave to the field modes a weight proportional to the phase space volume, and this was giving the high-frequency modes excessive weight. Probably he was also influenced by the then recent discovery by J. J. Thomson that electricity was not continuous, but electrons existed. He realized that the situation was sufficiently desperate to justify a bold conjecture.[12] He assumed that the energies of the oscillators cannot vary with continuity, but the only allowed values are

$$\epsilon_n = nh\nu, \tag{5.89}$$

with n a positive integer. This assumption fixed the theory and gave complete agreement with experiments by taking h$\sim 6.6269710^{-34}$ Joule $\times$ second. This is Planck's constant. Now, from the partition function

$$Z = \sum_n e^{-\frac{\epsilon_n}{K_B T}} = \frac{1}{1 - e^{\frac{h\nu}{K_B T}}} \tag{5.90}$$

and the average energy per oscillator turns out to be

$$u(\nu) = \frac{h\nu}{1 - e^{\frac{h\nu}{K_B T}}}. \tag{5.91}$$

This formula has a special role in Physics. Since 1900 no changes have affected it, but it has gone through all the revolutions unscathed and has been verified experimentally from radio waves to gamma rays, over 19 orders of magnitude; moreover it agrees with Relativity as well (on a Lorentz transformation, a black body remains black, although its temperature is modified). It has been the source of inspiration for many other important developments, some of which will be presented in the following.

In Chap. 25, we shall resume Planck's formula in the general framework of Quantum Statistical Mechanics.

5.10 Grand-Canonical Ensemble

In most problems, a more general ensemble is preferable. It is the Grand-canonical Ensemble in which the system $\mathcal{S}$ can exchange with the bath $\mathcal{B}$ besides energy, even particles. Instead of fixing the energy U = E as in the microcanonical ensemble, we

[12] Making conjectures about a hard problem is absolutely in line with the Galilean method; however, conjectures with no support from experiment and counterfactual assumptions are presented by some theoreticians as facts or even as great discoveries. This is quite bad.

fix the temperature T as in the canonical one; instead of fixing the number N of particles as one does in both the microcanonical and canonical scheme, one keeps the chemical potential μ fixed. Now all N values are possible, and the G space becomes the union of all those who describe the system with N defined. It is this monster space that must be divided into cells. In the search for the most probable distribution, we must maximize W at constant N, and this requires a new Lagrange multiplier that turns out to be μ. Ultimately, the formalism works as in the canonical scheme, but with the Hamiltonian $H(p, q)$ replaced with $H(p, q) - \mu$ N. Using the discrete notation, the grand-canonical partition function is

$$Z_G = \sum_i e^{-\beta[\epsilon_i - \mu N]}. \tag{5.92}$$

The grand-canonical average of any A becomes

$$\overline{A} = \frac{1}{Z_G} \sum e^{-\beta[\epsilon_i - \mu N]} A(i). \tag{5.93}$$

The grand-canonical internal energy is given by

$$U = -\frac{\partial}{\partial \beta} \log Z_G \tag{5.94}$$

in analogy with the canonical case. The only new quantity is the average number of molecules,

$$N = \frac{\partial}{\partial \mu \beta} \log Z_G. \tag{5.95}$$

5.11 Random Numbers and Monte Carlo Methods

5.11.1 Uses of Random Numbers

Many activities require a draw or the intervention of chance and the generation of random numbers. To quote a few:

Statistical analysis. Random sampling and simulation studies often require random numbers to create representative samples or simulate uncertain events.

Cryptographic Systems: Cryptography relies on the generation of random numbers for creating secure encryption keys and ensuring confidentiality in communication.

Gaming and Gambling: In games of chance like casino games or online gaming, random numbers are used to determine outcomes, ensuring fairness and unpredictability.

Computer Simulations: Random numbers are used in computer simulations in Physics, Biology, Economics and Engineering.

Security Protocols. Random numbers are crucial for generating session keys, nonce values.

Artificial Intelligence and Machine Learning also use random numbers for introducing randomness in optimization algorithms.

5.11.2 Generation of Random Numbers

Truly random numbers can be generated experimentally, using, for example, the nuclear decay of radioactive substances, but it is usually preferable to generate pseudo-random sequences (that are deterministic but are started by a seed, as the clock of a computer, whose results are practically unpredictable). The most commonly used is the one by Lehmer. One starts by choosing three large secret integers a, b, m. Let the integer r_1 be the random seed. More random numbers r_i are then generated by the formula

$$r_{i+1} = (ar_i + b) mod \quad m.$$

That is, r_{i+1} is the remainder of $\frac{(ar_i+b)}{m}$; in this way, the ratio $\frac{r_i}{m}$ are pseudo-random numbers between 0 and 1.

5.11.3 Some Applications in Science

Random number generation has important applications in Science, often dealing with the calculation of integrals. A simple example will convince you of the power of the methods based on random numbers.

Given a square of side a, we can inscribe in it a circle of radius a/2. Suppose one uses a random number generator to produce pairs (x, y) with $-a < x < a, -a < y < a$. The dot with coordinates (x, y) is inside the circle with probability $\pi/4$. This is actually one way to approximate π.

In a one-dimensional integral $\int_0^L dx f(x)$, the error due to a finite mesh goes like $\epsilon \sim \frac{\Delta L}{L}$. With $n \sim \frac{1}{\epsilon}$ points, the integral starts to converge, unless f has too narrow structures.

In d dimensions, the number of times one must compute f goes like $n \sim \epsilon^{\frac{1}{d}}$. If $d = 100$ and we want $\epsilon < \frac{1}{1000}$, we need 10^{300} computations of f. Assuming that one computation of f takes 10^{-6} s, the calculation lasts about 10^{294} s. Comparing this figure with the age of the universe, which is less than 10^{18} s, we may conclude that the waiting time is excessive and a faster method must be sought!

The Monte Carlo method is what we need. Actually, the variance of the Monte Carlo method decreases like $\frac{1}{\sqrt{N}}$ with increasing the number N of points, regardless of d. Moreover, if one knows which regions of the integration field contribute most, one can improve the convergence by an *importance sampling*. Molecular dynamics simulations can be based on the numerical solution of the equations of motion of samples of molecules; the samples are prepared using a number of realizations of the phenomenon under study, weighted by the probability of occurrence (the Metropolis method). In this way, one can compute mean values of the thermodynamic potentials.

The Monte Carlo approach is also applied with success to quantum mechanical simulations. The Car–Parrinello method is often used for this purpose. It is based on quantum mechanical *ab initio calculations* using the Density Functional approach (see Sect. 21.4), which is an efficient method to find the ground state density and energy in many-electron systems. Due to its computational complexity, it is applied on model systems that are much smaller than those that can be considered in classical molecular simulations. Car–Parrinello calculations have also been applied successfully to systems of interest for biology.

5.12 Irreversible Gates and Toffoli Gate

In 1854 at the University College of Cork, George Boole developed a new algebra to embody the logic of propositions. This is at the foundation of modern computing. The basic unit of information is the bit, which can be $0 \equiv$ FALSE or $1 \equiv$ TRUE. Logic gates are devices that transform some input bits into output bits according to some prescribed rule. For instance, NOT is a gate with one input bit and one output bit; it changes TRUE into FALSE and FALSE into TRUE. Some gates have two inputs and one output. A AND B (written as $A \wedge B$) gives 1 if both inputs are 1, 0 otherwise, while A OR B (written as $A \vee B$) gives 1 if at least one of the inputs is 1. NOR gives 1 if OR gives 0 and 0 if OR gives 1. Similarly, NAND gives the opposite reply to AND.

NAND and NOR are two examples of the so-called universal gates. This means that by combining enough NAND gates (or NOR gates), one can perform any logical operation and eventually carry out any calculation in Boolean logic algebra. One could set up a computer that only operates with combinations of one kind of gate, such as NAND or NOR. However, such gates have the property of wasting some information. For instance, AND gives 0 if the input is 00, 01 or 10, and NAND gives 1 in such cases; in this way, the output does not allow for the unique specification of the input. One classifies such gates as irreversible. In 1961, the German Rolf Landauer worked at IBM and realized that computation must be subject to the laws of Thermodynamics like any other macroscopic process. In particular, this implies that a minimum entropy amount equal to $K_B T \log 2$ is produced when a bit of information is erased. To avoid this, one can use reversible gates, which, in order to avoid information loss, must have as many output as input channels. NOT is a

trivial example. Many of them perform permutations of the input bits, and therefore can be represented by permutation matrices. For example, with three inputs, one has the 8 combinations 000, 001, 010, 011, 100, 101, 110, 111. Taking these in the above order, the Toffoli gate, proposed by Toffoli in 1980, is represented by the permutation matrix

$$\begin{pmatrix} 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&1&0 \end{pmatrix}.$$

This gate gives an output identical to the input in the following cases: (0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1). Instead, (1, 1, 0) gives (1, 1, 1) and (1, 1, 1) yields (1, 1, 0). The Toffoli gate is clearly reversible and can be shown[13] to be universal.

Reference

Williams, C.P.: Explorations in Quantum Computing. Springer (2011)

[13] Williams (2011).

Chapter 6
Special Relativity

Everyone is entitled to use their own measurements to describe physical reality, and the descriptions of observers who are in different inertial frames are related by simple transformation laws. This was well understood by Galileo, but the finite speed of light required Einstein to fit electromagnetism into the scheme, with paradoxical and astounding consequences.

6.1 Galileo's Ship

Galileo stated the *Principle of Relativity* using a series of thought experiments to be performed inside a ship. This was revolutionary in a time dominated by Aristotelian misconceptions, when the common wisdom was that there is an absolute rest system and a force is needed to keep objects moving.[1] The worst difficulty was that the 2000-years-old teaching of Aristotle was upheld by the Church.

It was the first time that thought experiments were introduced and their power demonstrated, and many ways, that was the beginning of Science in the modern conception of the term. Galileo's *Dialogo dei Massimi Sistemi* is also a masterpiece of Italian Literature, written in the colorful but still perfectly clear Italian of 1600. He demonstrated that no experiment conducted inside the ship without looking outside could reveal the state of motion of the ship, as long as it is uniform. Let me summarize three novel principles that were clearly demonstrated by Galileo in that work and in other publications:

[1] However, the times were ripe and some of the works by Giordano Bruno (1548–1600) imply correct ideas about the relativity of motion.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_6

1. There is *inertial frame of reference*, such that a free body moves along a straight line with constant speed once the friction is removed;
2. Given an inertial frame, all the frames that move with constant speed in it (without rotations) are also inertial;
3. No experiment can distinguish one inertial frame from another, and the laws of Physics are the same for all.

Remarkably, he arrived at that without ever having access to any even approximately inertial system (apart from a freely falling one), since the ship actually is in a field of gravity. Let us discuss (in modern terms) the implications (Fig. 6.1).

Consider two inertial references $\mathcal{K}$ and $\mathcal{K}'$ and a material point P; let the coordinates of P be $\overrightarrow{r} = (x, y, z)$ in $\mathcal{K}$ and $\overrightarrow{r}' = (x', y', z')$ in $\mathcal{K}'$. If the origin of $\mathcal{K}'$ travels in $\mathcal{K}$ with speed $\overrightarrow{u}$, and the two origins coincide for $t = 0$, then $\overrightarrow{r}'$ is related to $\overrightarrow{r}$ by

$$\overrightarrow{r}' = \overrightarrow{r} - \overrightarrow{u}\, t. \tag{6.1}$$

The time is just a parameter, the same in both references, and Newton's equations are trivially the same; a change of variables in the Lagrangian makes the change. This is the Galilean Relativity. In the nineteenth century everyone believed in Classical Mechanics, but the crisis arrived with Maxwell's equations, which predict that light moves with velocity c. Velocity relative to what reference? For a long time, it was thought that the privileged reference in which Maxwell's equations hold must

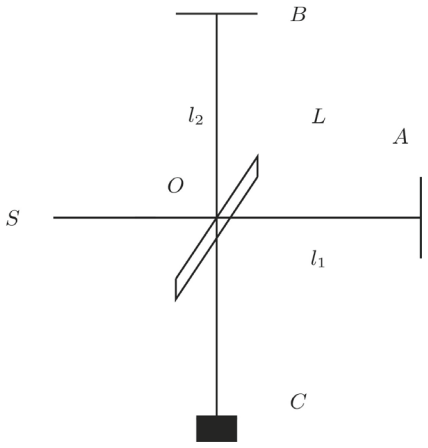


Fig. 6.1 Scheme of the Michelson interferometer used to seek a dependence of c on the direction. Light coming from a source S is split into O by a half-silvered mirror L making an angle of 45°; the rays OA and OB are reflected by mirrors that send them back to L. Here, the ray BO is split into OC and OS, while AO is split into OS and OC. In C there is an eyepiece, and one can see the interference between the rays SOAOC and SOBOC. Any difference of the light speed in the two arms should produce a phase shift. No dependence on the direction was found, despite the fact that the Earth travels at 29 km/s around the Sun

be sought by seeking variations of the speed of light with direction. A celebrated experiment used the Michelson interferometer of Fig. 6.1 but failed to find any effect. Nowadays, interferometric experiments using laser light measure the speed of light to a part in 10^9 or better, and any discrepancy would be a celebrated discovery, yet none has been reported. So we start to discuss Einstein's Special Relativity, which is based on the simple assumption that Galileo's Relativity is not restricted to Mechanics, but all the Laws of Physics must hold in the same form in any inertial system. In General Relativity, all reference systems are included, but this is deferred to Chap. 8.

Before closing this introduction, I wish to stress the enormous practical importance of Relativity in an amazing variety of applications. Our technology is based on calculations, not on trial and error; the calculations use condensed matter theory, and the modern condensed matter theory is necessarily relativistic, otherwise the calculations turn out wrong. Molecular calculations require relativistic quantum theory, and have applications in all fields. Accelerators are used in medicine and are based on Relativity. Also, GPS uses special and general Relativity. In this way, the speculative ideas of Galileo and Einstein have been world-changing.

6.2 From Thought Experiments to the Lorentz Transformation, Using Einstein's Undeclared Assumption

Einstein[2] started from the idea that the Maxwell equations were right and the laws of Electromagnetism must hold in any inertial system. In particular, any measurement of c yields the same c, with no dependence on the speed of the source of the light and no privileged reference system. Einstein in his Special Relativity Principle made the stronger statement that all the laws of Physics must have the same form in all inertial reference systems. This was in line with Galileo, but the Principle of General Relativity states that all the laws of Physics must have the same form in all reference frames, including the accelerated ones. In this chapter we develop the special theory (Fig. 6.2).

An event, like the emission or the absorption of a light signal at some point in some instant, is something real, something that will receive different space–time coordinates by different observers, but is something objective in itself. All phenomena consist of a succession of events, each characterized in a given frame by 4 coordinates $(\vec{x}, t)$. We must give precise meanings to these coordinates. We need to examine critically the physical (i.e., operational) meaning of the notions of space separation and time interval between two events. The theory must refer to real facts

[2] Albert Einstein (Ulm 1879–Princeton 1955) is so famous that it is hardly necessary to mention his merits. Definitely, the most important physicist since the time of Newton, besides inventing and developing Special and General Relativity he gave a formidable contribution to Quantum Mechanics as well. Being of Jewish origin, he was forced to flee from Germany already in 1933. He spent the rest of his life in Princeton (USA).

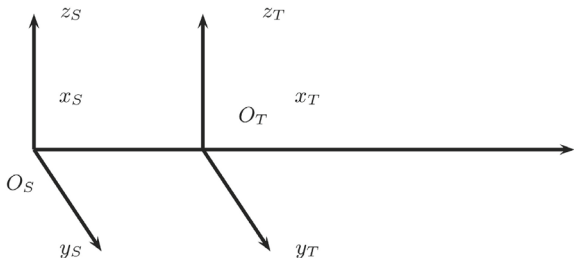


Fig. 6.2 The thought experiments comparing the results of measurements performed on the train with those performed in the station. The train reference travels with speed u along the x axis relative to the station; the train crosses the station for $t_S = t_T = 0$

and measurable quantities that likewise will be transformable from one reference to the other. As we shall see shortly, all such objects will be described mathematically as tensors, which are generalized vectors.

Meters, Clocks and Frames

Each observer is assumed to set up his reference system, including a Cartesian reference, and to be able to set meters and clocks wherever they are needed. By a meter, one can measure the distance between two points provided that they are fixed; a clock measures the time between two events provided that they occur both at the clock. We need to specify an ideal procedure enabling the observer, at least in principle, to synchronize all his distant clocks. This can be achieved by sending a signal from the origin; each clock is set at the time $t = \frac{r}{c}$, where r is the measured distance from the origin. Then, the observer is ready to measure the space–time coordinates (x, y, z, t) of any event and give a sharp meaning to his measurements. The length of a moving rod is measured as the distance between two simultaneous events in which the ends are observed at the same t. The time interval between events that occur at distant locations is given by the difference of the times at the synchronized clocks. The meaning of physical quantities can only be defined by specifying how we measure them.

Next, we need to understand the relation between the measurements performed by different observers, and to be able to switch reference. Some quantities are invariant in a trivial way since they are defined in terms of measurements done in special reference systems. A stationary rod can be measured using a meter; such is a *proper length*, since all observers rely on the measurement done in a particular reference. Likewise, when two events in a given system occur at the same point, the time separation is a *proper time*. In the same way, the *rest mass* of a body is measured in a reference where it is motionless.

Instead of the ship used by Galileo, we shall use the example of a train that moves with constant speed $\overrightarrow{u}$ relative to a station. The train represents the reference (x_T, y_T, x_T, t_T), and the station (x_S, y_S, z_S, t_S).

The train driver and the station manager agree to make some measurements and compare the results. They agree on the relative velocity, although each of them

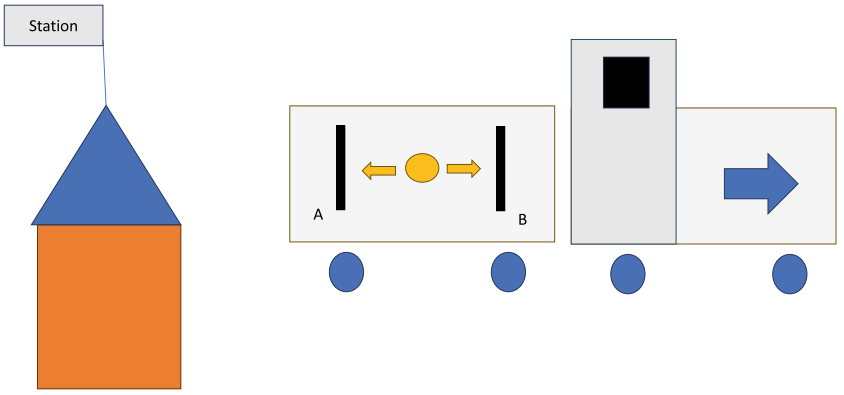


Fig. 6.3 Thought experiment showing the relativity of simultaneity

pretends to be motionless while the other is moving. Often they disagree on the results of the measurements. Each observer will interpret the experiments based solely on his own measurements. The relation between the results of both observers is fixed by the Principle of Relativity. The station manager can measure the distance between the rods of the rail using a meter. This is a proper length for him. But the train driver is also able to do the same, with his own meter, because although the iron of the rods is moving, the rods always appear to be the same for him. The conclusion is that the lengths in transverse directions (orthogonal to the velocity) are the same for both. This is actually an undeclared assumption made by Einstein in his landmark paper [Ann. Physik 322 (10), 891 (1905)] in which he introduced the Special Theory of Relativity (STR).[3]

6.2.1 Relativity of Simultaneity

The train driver switches a lamp at O and measures the time the light takes to reach mirrors A and B placed on the train at equal distances from O. To do so, he fixes a clock close to each mirror. The train driver finds that the light hits both mirrors at the same time (Fig. 6.3). The two hits are simultaneous events, taking place at different places but at the same time.

The line AOB is parallel to the velocity of the train and B is ahead. If the station manager wants to measure the arrival times at A and B, the experimental setup must include two clocks fixed to the ground and close to the mirrors A and B at the times when each receives the light, since the clocks must be in the reference of the

[3] See the paper in Research Article Open Access Volume 2 Issue 1 Journal of Physics and Optics Sciences "The Clock Riddle and Einstein's Third Postulate of Special Relativity", by Robert J. Buenker.

station. This arrangement is more complicated than the measurement on the train, but can be done. He finds that the light hits A first. This is in line with the fact that the speed of light is c, since during the propagation of the light, A approaches the point from which the light was emitted while B goes farther. Who is right? Both! In fact, an experiment is always right! Moreover, the simultaneity becomes absolute if the A-B distance vanishes. Finally, if the train driver sets mirrors at the same distance from O in a direction orthogonal to the rail, then the station manager agrees with him that the light reaches both mirrors at the same time; for both observers, the setup is symmetrical.

6.2.2 *Relativity of Time*

The train driver installs a lamp fixed to the ceiling at height h over the ground. When he switches the light on, how long does it take before it reaches the ground? For the driver,

$$t_0 = \frac{h}{c},$$

of course. This is a proper time, because nothing moves in the experiment, apart from the light.

The station manager agrees that the height is h, since it is a transverse length, but he finds that the light must cover the distance $\sqrt{h^2 + (ut_s)^2}$, where u is the speed of the train; hence, the time is given by $t_s = \frac{\sqrt{h^2+(ut_s)^2}}{c}$. This implies

$$t_s = \frac{h}{c} \frac{1}{\sqrt{1 - \left(\frac{u}{c}\right)^2}}. \tag{6.2}$$

For the station manager, the time is longer, because the distance covered by the light is longer. Actually, the time measured by the train driver is the shortest possible. It is a proper time, measured solely by using clocks. This is a general phenomenon. The time dilation

$$\tau = \frac{\tau_0}{\sqrt{1 - \frac{u^2}{c^2}}},$$

where τ_0 represents a proper time, is the reason why a relativistic muon lives much longer in the laboratory than in the rest system.

But one should not think that "in reality the experiment of the train driver lasts $2\frac{L}{c}$ and it *appears* to be longer to the station manager." There is no privileged reference, and the duration is really longer in the station; for instance, the station manager has a longer time to intervene in the experiment.

The station manager sees the clock in the train showing the time τ_0 while the station clock shows a longer time τ, and concludes that the train clock runs slow. Many people are tempted to think that the train driver then must see that the station clock runs too fast. By the principle of Relativity the train driver must see the station clock going slow. He says: I am motionless, while the station runs with a speed $-u$, so there is a time dilation by a factor of $\sqrt{1-\frac{u^2}{c^2}}$. Is that contradictory? Not at all. Both observers in 1 s can see what the other does in $\sqrt{1-\frac{u^2}{c^2}}$ s.

It is time to define the often-used symbol

$$\gamma = \frac{1}{\sqrt{1-\frac{u^2}{c^2}}} \tag{6.3}$$

and express the time dilation by $\tau = \tau_0\gamma$.

It is common practice also to introduce also the symbol

$$\beta = \frac{u}{c}. \tag{6.4}$$

6.2.3 *Relativity of Lengths*

In the next experiment, the station manager and the train driver agree to use the Michelson interferometer with SA parallel to the velocity $\overrightarrow{u}$ of the train and O at the origin of his reference; the station manager puts a lamp, a light detector and a clock at O. His report on the experiment is as follows.

With the train traveling at velocity u, at time t=0, the lamp emits a flash of light; since the arms have equal lengths, the light reaches both mirrors A and B simultaneously at time $t_1 = \frac{L}{c}$ *and both reflected signals return to O at time* $t_2 = \frac{2L}{c}$.

The simultaneity of the two returns to the origin is absolute (same time and same place). The experiment can be arranged such that the first flash starts from the origin at time 0 also in the reference of the station.

The reflection from A has space–time coordinates (x_T, y_T, z_T, t_T) with $t_T = (L, 0, 0, \frac{L}{c})$ and the reflection from B coordinates $(0, L, 0, \frac{L}{c})$; at time $(t_{OBO})_T = (t_{OAO})_T = 2\frac{L}{c}$, the two flashes return simultaneously in O.

The station manager agrees that mirror B is at $y = L$ since this is a transverse length, but considers this experiment a way to measure the length of OA not by observing O and A simultaneously but by observing them at different times. Since OB moves with velocity u, the ray OBO becomes the two equal sides of a triangle having as its basis the length of the distance covered by O during the time τ_{OBO}. Letting O' and O" denote the initial and final positions of O, $O'O'' = u\tau_{OBO}$, and the length covered by the ray is

$$O'BO'' = 2\sqrt{L_2^2 + \left(\frac{u\tau_{OBO}}{2}\right)^2} = \sqrt{4L_2^2 + (u\tau_{OBO})^2}.$$

The time measured in the reference of the station is longer, since the light makes a longer trip; actually,

$$\tau_{OBO} = \frac{2L_y}{c}\frac{1}{\sqrt{1-\frac{u^2}{c^2}}}, \tag{6.5}$$

but this is just the time dilation (6.2) noted above. On the other hand, the station manager cannot agree on the length of the arm OA parallel to the velocity. Let us find $L_S = L_{OA}$ in the reference of the station. During the time τ_{OA} of the OA trip, the mirror A initially at L_{OA} has moved forward by a distance $u\tau_{OA}$. The actual distance covered is $L_{OA} + u\tau_{OA}$. Since $\frac{L_{OA}+u\tau_{OA}}{\tau_{OA}} = c$, $\tau_{OA} = \frac{L}{c-u}$. Similarly, the reverse trip takes the time $\tau_{AO} = \frac{L}{c+u}$. To sum up, the time taken to go to A and back is

$$\tau_{OAO} = \frac{2L}{c}\frac{1}{1-\frac{u^2}{c^2}}.$$

Since it is an absolute fact that $\tau_{OAO} = \tau_{OBO}$, it follows that

$$L_S = L\sqrt{1-\frac{u^2}{c^2}}. \tag{6.6}$$

The proper length (measured by a meter at rest) is the longest. A shortening by 1% requires $u = 0.15c$, and the effect is too small to be noted in everyday life. Of course, the descriptions of both observers are equally right.

Note that the shortening of moving bars and the time dilation are *real* effects. Such are the results of any measurement.

6.2.3.1 The Twin Paradox

Anthony and Bartholomew are twins born in an inertial system $\mathcal{S}$. Bartholomew starts at time t_1 on a spacecraft, reaches a star 12 light years away traveling at $0.6c$ and soon returns home with the same speed. Anthony stays at home and celebrates the return of his brother at time t_2 after $t_2 - t_1 = 2 \times \frac{12}{0.6} = 40$ years. Meanwhile Bartholomew is aged $t_2' - t_1' = \int_{t_1}^{t_2} \mathrm{dt}\sqrt{1-(0.6)^2} = 40\times\ 0.8 = 32$ years, and is younger than his twin!

Bartholomew says: *I never traveled, it was Anthony who has taken a curved path with speed* $0.6c$. *Also, during the trip, I have often observed that terrestrial clocks were slow. Really, then I would have expected to find Anthony younger than myself.*

This is the paradox. The resolution is that Bartholomew must have felt inertial forces during acceleration at the start, when he bent to return home and during the

braking at arrival. So he has no right to consider himself stationary in an inertial system. The frame of Antony, instead, was inertial.

This experiment was conducted in 1972 comparing two Cesium clocks with each other, one of which had toured the world in an airplane. Despite the smallness of the velocity compared to c, the relativistic effect was found as Einstein had predicted.

6.2.4 *Relativity of Synchronization*

Two clocks at rest in the same inertial reference can be synchronized in the following obvious way. The distance L is a proper distance and in principle is measured by using a meter, after which one of the clocks, say clock A, emits a signal at time t_A and the other is set at time $t_A + \frac{L}{c}$ when the signal arrives. The observers in different inertial systems, however, do not agree that the clocks are synchronized. An observer who travels on a spacecraft from A to B with speed u relative to the clocks and is at A at time t_A finds that the distance is actually $\frac{L}{\gamma}$, so he reaches B at the time $t_A + \frac{L}{u\gamma}$; this is his proper time. Instead, clock B indicates $t_A + \frac{L}{u}$. So, in the system of the spacecraft the clocks are not synchronized.[4] The synchronization is relative to the reference.

6.2.5 *More Striking Paradoxes*

Suppose a spaceship and a tunnel in the space station have the same proper length L; the spaceship passes at relativistic velocity through the tunnel. For the tunnel reference, due to the relativistic contraction, the spaceship is shorter than the tunnel; therefore there is a time interval during which the spaceship is traveling inside the tunnel; in other terms, the gates can be shut down with the spaceship inside the tunnel. For the crew of the spaceship, however, that is absurd since the tunnel is shorter than the space shift. This is in line with the relativity of time noted above. The station keeper finds that the back of the space ship enters the tunnel before the front exits, while the crew sees the opposite order of the events. However, the paradox is not entirely over, because some other chapters of Theoretical Physics require an absolute arrow of time, like the fact that entropy cannot decrease (Chap. 5).[5]

---

[4] If the astronaut is unaware of Relativity, he will find a still stronger disagreement. He notes that the clocks in the Earth system go slow by a factor of $\frac{1}{\gamma}$; so, if he keeps the length $\frac{L}{\gamma}$ that he measures as the true length, then he expects that on the Earth, the elapsed time should be $\frac{L}{u\gamma^2}$, while in reality it is $\frac{L}{u}$, as we know.

[5] See the very interesting paper by Avshalom C Elitzur and Shahar Dolev, https://researchgate.net/publication/226976724.

6.2.6 The Lorentz Transformation

We were able to deduce some of the most characteristic effects of Special Relativity through thought experiments, using only the fact that c is the same in all the inertial systems. To make progress, we must derive the Lorentz transformation. This can be done in a formal, abstract way, but we are ready to determine the result very simply through a new thought experiment, again with the help of the station manager and the train driver. We arrange that their origins O_S and O_T coincide at time $t_S = t_T = 0$. The station manager prepares a traffic light (which is fixed in the station reference $\mathcal{K}_S$) at position x_S along the rail; a timer turns the light on at some station time t_S. This is an event E with coordinates (x_S, t_S).

In the reference $\mathcal{K}_T$ running at speed u, it is observed that the event E has coordinates $(x_T^{(E)}, y_T^{(E)}, z_T^{(E)}, t_T^{(E)})$. As we know, the y and z coordinates are the same for both references and we can forget about them. The train driver measures the distance $(O_S E)_T$ of E from the station, which is in O_S. The distance along the x axis of E from O_T is just $x_T^{(E)}$. In order to find the distance from O_S, he adds the distance $O_S O_T = ut_T$ from the station, still measured on the train by a meter and a clock; the result is evidently

$$(O_S E)_T = x_T^{(E)} + ut_T^{(E)}.$$

Instead, the station manager measures $x_S = (O_S E)_S$ by a meter, as a proper length. Then, taking into account the relativistic contraction,

$$(O_S E)_T = x_S \sqrt{1 - \left(\frac{u}{c}\right)^2} \equiv \frac{x_S}{\gamma};$$

therefore, (understanding $^{(E)}$),

$$x_S = \gamma[x_T + ut_T]. \tag{6.7}$$

By the principle of Relativity, it follows that

$$x_T = \frac{x_S - ut_S}{\sqrt{1 - \frac{u^2}{c^2}}} \equiv \gamma[x_S - ut_S]. \tag{6.8}$$

Eliminating x_T from Eqs. (6.7) and (6.8), one obtains $x_S = \gamma^2(x_s - ut_S) + \gamma ut_T$; but $1 - \gamma^2 = -\beta^2\gamma^2$. In this way one can find t_T in terms of measurements done by the station manager:

$$t_T = \gamma[t_S - \frac{u}{c^2}x_S]. \tag{6.9}$$

Having completed the Lorentz transformation, (6.8) and (6.9), we realize that it is symmetric in x and ct. The Galileo transformation (6.1) is the limit for $c \to \infty$.

The inverse transformation is

$$x_S = \gamma\,[x_T + ut_T]\,,\ ct_S = \gamma\left[ct_T + \frac{ux_T}{c}\right], \tag{6.10}$$

with $y_S = y_T, z_S = z_T$. Usually, one replaces the time with a zeroth space dimension $x_0 = ct$ and the velocity by the pure number $\beta = \frac{u}{c}$; then the transformation becomes

$$\begin{cases} x_T = \gamma(x_S - \beta x_{0S}), \\ x_{0T} = \gamma(x_{0S} - \beta x_S), \end{cases} \tag{6.11}$$

still with $y_T = y_S$ and $z_T = z_S$. The Lorentz transformation leads to the phenomenon of the Lorentz contraction, whereby a bar moving parallel to its length is shorter than its rest length. Suppose the station manager wishes to measure the length of a bar on the train parallel to the velocity of the train. The train driver can measure L using a meter. If one uses (6.10) for the ends of the bar setting the same t_T for both, the result is that the bar length increases by a factor γ. This is wrong, because the positions of both ends A and B must be taken simultaneously in the station, not in the train. In order to find the correct condition, we apply (6.11) to both ends of the bar, then write the difference, $\Delta x_T = \gamma\Delta x_S - \beta\Delta x_{0S}$, $\Delta x_{0T} = \gamma(\Delta x_S - \Delta x_{0S})$; now we set $\Delta x_{0S} = 0$. Since $\Delta x_S = \gamma\Delta x_S$, one finds that $L_T = \gamma L_S$. This is the contraction $x_B^{(S)} - x_A^{(S)} = L\sqrt{1 - \frac{u^2}{c^2}}$.

Problem 6.1 The space–time coordinates of two events as measured by an observer O are $x_1 = 6 * 10^4$ m, $y_1 = z_1 = 0$, $t_1 = 2 * 10^{-4}$ s, $x_2 = 12 * 10^4$ m, $y_2 = z_2 = 0$, $t_2 = 1 * 10^{-4}$ s. The observer O' travels with a speed v relative to O along the x axis, and finds that the two events are simultaneous. Find v using $c = 3 \times 10^8\ \frac{\text{m}}{\text{s}}$.

Solution 6.1 Since $ct_1 - \frac{u}{c}x_1 = ct_2 - \frac{u}{c}x_2$, one finds that $\frac{v}{c} = -\frac{1}{2}$.

Problem 6.2 The rocket O' travels with speed v relative to a space station O. The astronaut looks at the station through a 4 m porthole and says that the size of the porthole is the same as the size of the window in the station. The station manager using a meter finds that the window is 5 m long. Evaluate v.

Solution 6.2 From $5 = 4\gamma$ one finds $\frac{v}{c} = \frac{3}{5}$.

6.2.7 Apparent Superluminal Motion and Other Spectacular Effects

Suppose an alien rocket L meters long is approaching the Earth at a speed u. As we have seen in Eq. (6.6), its length is contracted by a factor $\sqrt{1 - \frac{u^2}{c^2}}$. Actually, however, we should see it lengthened, not contracted. The reason is that the length in Eq. (6.6) refers to the distance between the simultaneous positions of the ends of the rocket in a reference where the Earth stands still. However, when the astronomers on Earth stare at the rocket, they see an elongated image because the signal from the tail

of the rocket was emitted earlier and the signal from the tip was emitted later, so that both arrive at the telescope simultaneously. For the same reason, a rocket that goes away must appear shortened more than expected according to Eq. (6.6) because the signal from the tip must be emitted earlier than the signal from the tail. This sort of effects was noted first by the Austrian physicist A. Lampa.[6]

Similar considerations explain the apparent superluminal motion first observed in 1973, in the quasar 3C279, which has a jet moving (apparently) at 10c. Several other examples have been discovered since. Of course, estimates of velocities depend on estimates of distances based on red shifts, and give the component of the velocity which is orthogonal to the line of sight. However, the qualitative explanation is the same as in the case of the approaching rocket: due to the finite speed of light, approaching jets appear much lengthened.

Interestingly, a combination of the relativistic contraction and the finite velocity of light changes the observed shape of all objects, except spheres. In the case of a spherical body, the length contraction tending to flatten the aspect is compensated by the lengthening due to the finite speed and a sphere remains a sphere, All other shapes are altered, however, and at high speeds one becomes able to observe part of the back of an incoming body.

For more information, computer simulations and and pictures, see https://www.spacetimetravel.org/bewegung. Spectacular images can be seen in https://www.tempolimit-lichtgeschwindigkeit.de (in German).

6.2.8 *Constant Acceleration, Rindler Metric and Interstellar Missions*

Consider a space ship starting from the rest along the x axis ($x = 0$ for $t = 0$, both measured in the station) from an inertial space station in such a way that the crew feels a constant acceleration a, like an artificial gravity. This concept will be developed in Chap. 8. The transverse coordinates can be ignored. The inertial system has space–time coordinates (x, t) while those used by the crew are $(x', t' \equiv \tau)$. We must find the motion of the space ship as observed by the station master, who measures the velocity $V(t)$.

In the space ship at any time the crew agrees with the station master about the relative velocity V, but feels a gravity. Its intensity is $a = \frac{dV}{d\tau}$. Consider an infinitesimal interval; in the ship the clocks run and we may write $ds^2 = c^2 d\tau^2$; in the station system, this corresponds to the clocks telling the time dt and the ship moving by dx.

[6] Lampa (1924).

Since the interval s is invariant,

$$ds^2 = c^2 d\tau^2 = c^2 dt^2 - dx^2 = dt^2(c^2 - V^2) \Longrightarrow d\tau = \frac{dt}{\gamma}. \tag{6.12}$$

Clocks run slower in the space ship. So, the time measured by the crew if $\tau = 0$ for $t = 0$ is

$$\tau = \int_0^t dt \sqrt{1 - \frac{V(t)^2}{c^2}} < t. \tag{6.13}$$

In Chap. 8, when dealing with gravity, we shall find that clocks run slow in a gravity field. Indeed, as we shall see, an accelerated reference works locally like a gravity field. The present result is an anticipation of the gravitational red shift of General Relativity.

We must still find $V(t)$, but putting $d\tau = \frac{dt}{\gamma}$ into $a = \frac{dV}{d\tau}$ we find

$$\frac{d}{dt} \frac{V(t)}{\sqrt{1 - \frac{V(t)^2}{c^2}}} = a. \tag{6.14}$$

Integrating with the initial condition $V(0) = 0$, one finds

$$V(t) = \frac{at}{\sqrt{1 + (\frac{at}{c})^2}}. \tag{6.15}$$

For long times, this approaches c; we can integrate again, and setting $x(0) = 0$,

$$x(t) = \frac{c^2}{a} \left(\sqrt{\left(1 + \left(\frac{at}{c}\right)^2\right)} - 1 \right). \tag{6.16}$$

By inverting this relation, one finds the station time versus the distance x:

$$t(x) = \sqrt{\left(\frac{x}{c}\right)^2 + \frac{2x}{a}}. \tag{6.17}$$

Moreover, inserting in Eq. (6.13) gives the proper time versus t,

$$\tau = \int_0^t d\tau = \int_0^t dt \sqrt{1 - \frac{V^2}{c^2}} = \frac{c}{a} \text{arcsinh} \left(\frac{at}{c}\right). \tag{6.18}$$

Solving for t, one finds the station time as a function of the crew proper time:

$$t(\tau) = \frac{c}{a} \sinh \left(\frac{a\tau}{c}\right). \tag{6.19}$$

Moreover,

$$x(\tau) = \frac{c^2}{a}(\cosh(x) - 1). \tag{6.20}$$

Finally,

$$v(\tau) = c \tanh\left(\frac{a\tau}{c}\right). \tag{6.21}$$

6.2.8.1 Mission to Nearby Stars

Suppose we are able to build a star ship capable of a constant acceleration a of the order of the Earth's gravitational acceleration for years. At about half way, the Captain orders the ship to capsize, so the acceleration becomes $-a$ and the trip ends at zero speed (I am neglecting the proper motion of the star). We need some numbers: one year $= 3.154\ 10^7$ s, 1 light year = $9.46\ 10^{15}$ m and $c = 3 * 10^8$ m s^{-2}. Then, the time in years taken for the crew to reach the destination t_{Earth} measured in the station and τ_{crue} measured by the astronauts is reported in the following Table.

Star	D (light years)	t_{Earth}	τ_{crue}
Proxima Centauri	4.24	5.85	3.5
Altair	14	15.8	5.35
Vega	25	26.88	6.36
Arcturus	36,7	38.6	6.95

We are far from developing the technology needed for this sort of timetables and at the velocities involved even a small stone is a disaster. On the other hand, there are 103 main sequence stars within 20 light years; if an advanced civilization exists, the idea that aliens are paying visits does not seem unreasonable. Some surprising consequences of these results will be developed in Chap. 25.

6.2.9 Relativistic Addition of Velocities

While the train moves along the x axis with speed u relative to the station, the train driver sets up a reference with coordinates x_T, t_T. Then, he observes a body moving along the train at a speed

$$W = \frac{dx_T}{dt_T}.$$

For the station manager, who sets up a reference with coordinates x_S, t_S as usual, the speed is $V = \frac{dx_S}{dt_S}$, where

$$dx_S = \gamma[dx_T + udt_T],$$

$$cdt_S = \gamma[\frac{u}{c}dx_T + cdt_T].$$

So,

$$V = \frac{W + u}{1 + \frac{uW}{c^2}}. \tag{6.22}$$

For W and u less than c, V is always $< c$.

More generally, suppose that the body moves relative to the train with a velocity W along a straight line making an angle θ_T with the x_T axis:

$$x_T = Wt_T \cos(\theta_T), \quad y_T = Wt_T \sin(\theta_T), \quad z_T = 0. \tag{6.23}$$

The station manager will observe the body moving with the trajectory

$$x_S = Vt_S \cos(\theta_S), \quad y_T = Vt_S \sin(\theta_S), \quad z_S = 0. \tag{6.24}$$

Substituting into the Lorentz transformation, one finds that

$$V^2 = \frac{W^2 + u^2 + 2Wu\cos\theta_T - (\frac{W^2u^2}{c^2})\sin^2(\theta_T)}{[1 + (\frac{Wu}{c^2})\cos(\theta_T)]^2}, \tag{6.25}$$

with

$$\tan(\theta_S) = \frac{W\sin(\theta_T)\sqrt{1-\beta^2}}{W\cos(\theta_T) + u}. \tag{6.26}$$

If W is parallel to the velocity of the train, we recover the result (6.22).

Problem 6.3 A rocket R flies away from Earth E along a straight line, while a UFO is seen from E and from R. Both objects proceed in the same direction. An observer on E finds that the UFO is going at $0.5c$, while the pilot of the rocket finds that it goes at $-0.5c$. What is the speed of R relative to E?

Solution 6.3 $\frac{v}{c} = 0.8$.

Problem 6.4 In the reference (x_S, t_S), the station manager sees a train and a rocket traveling in the same direction (say, the x direction), and finds that the speed of the train is u. The train driver in his reference (x_T, t_T) finds that the speed of the rocket is v. Write the Lorentz transformations from the station to the train and from the train to the rocket. Then, eliminating (x_T, t_T) obtain the transformation from the station to the rocket. Show that this is of the Lorentz form and verify the theorem of the sum of velocities.

Solution 6.4 The required transformation reads as

$$x_S = \gamma(u)\gamma(v)[(1+uv)x_R + (u+v)t_R], t_S = \gamma(u)\gamma(v)[(1+uv)t_R + (u+v)x_R].$$

This is a Lorentz transformation provided that $\gamma(u)\gamma(v)(1+uv) = \gamma(w) = \frac{1}{\sqrt{1-w^2}}$, where w is the speed of the rocket in the reference of the station. Solving, one finds that $w = \frac{u+v}{1+uv}$.

6.3 The Geometry of Special Relativity: Minkowsky Chronotope

The Lorentz transformation mixes time and space coordinates. The transverse lengths are conserved, but the lengths along the velocity are not, so distances depend on the reference. However, using Eqs. (6.8), (6.9), and dropping the subscript S,

$$c^2t_T^2 - x_T^2 = \frac{c^2t^2 - 2ux\ \mathrm{t} + \frac{u^2x^2}{c^2}}{1 - \frac{u^2}{c^2}} - \frac{x^2 - 2\ \mathrm{u}xt + u^2t^2}{1 - \frac{u^2}{c^2}} = c^2t^2 - x^2.$$

The invariant quantity is the *interval* s, defined[7] by its square

$$s^2 = -r^2 + c^2t^2. \tag{6.27}$$

The interval between two events, unlike a distance in a four-dimensional Euclidean space $\mathbb{R}^4$, can be positive, zero or negative; a vanishing interval $s^2 = 0$ is called *light-like*, because the light can reach one event from the other. When $s^2 < 0$, the interval is *space-like* and imaginary, while if $s^2 > 0$ the interval is *time-like*. Space–time is a four-dimensional space, the Chronotope, whose coordinates $x_0 = ct$, $x_1 = x$, $x_2 = y$, $x_3 = z$ are lengths. A point in the Chronotope is called an event. In a four-dimensional Cartesian space $\mathbb{R}^4$, any linear transformation that leaves square distances $\sum_{i=1}^4 x_i^2$ invariant is a rotation (improper, if it involves a reflection). In the Chronotope instead, we can make Lorentz transformations with invariant s^2. Such a space is called pseudo-Euclidean. In $\mathbb{R}^4$, when the distance between two points vanishes, the two points coincide, while a light-like interval means that a signal can travel from one event to the other.

We can keep the Euclidean sum of the squares formula if we are ready to use the formal trick of replacing x_0 with $x_4 = ix_0 = ict$. The Minkowsky Chronotope is the pseudo-Euclidean space where a point has coordinates:

$$x_1 = x, x_2 = y, x_3 = z, x_4 = ict.$$

The Lorentz transformation becomes

$$\begin{cases} x_1' = \gamma(x_1 + i\beta x_4), \\ x_2' = x_2, \\ x_3' = x_3, \\ x_4' = \gamma(x_4 - i\beta x_1). \end{cases} \tag{6.28}$$

Any object $v = (v_1, v_2, v_3, v_4)$ with 4 components that Lorentz-transform like x_1, x_2, x_3, x_4 is called a four-vector. We shall find several physical observables that

[7] About half of the authors define it as $s^2 = r^2 - c^2t^2$.

behave as four-vectors. For instance, consider a particle moving with speed $v = \frac{dx}{dt}$ in some reference. The derivative $\frac{dx_\mu}{dt}$ has 4 components, but is no four-vector, because dt depends on the reference. One obtains a four-vector by differentiating the position four-vector x_μ with respect to invariant proper time of the particle[8] $d\tau = \frac{ds}{c}$. One finds the velocity four-vector $v_\mu = \frac{\delta x_\mu}{d\tau}$ with components $v_1 = \frac{dx}{d\tau}$, $v_2 = \frac{dy}{d\tau}$, $x_3 = \frac{dz}{d\tau}$, $i\frac{dt}{d\tau}$. This is called four-velocity. Since $\frac{dt}{d\tau} = \gamma$ with $\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$,

$$v_\mu = \gamma(\boldsymbol{v}, c), \tag{6.29}$$

where $\boldsymbol{v}$ is the three-dimensional Galilean speed.

The components are conventionally denoted by a Greek subscript, like x_α, with $\alpha = 1, 2, 3, 4$. The Latin indices, like i in $x_i, i = 1, 2, 3$, are used for the first 3 (space) components of the four-vector.

The special transformation (6.28) can be put in the convenient matrix form

$$x'_\mu = \sum_{\nu=1}^{4} \Lambda_{\mu\nu} x_\nu, \tag{6.30}$$

with

$$\Lambda = \begin{pmatrix} \gamma & 0\,0 & i\gamma\beta \\ 0 & 1\,0 & 0 \\ 0 & 0\,1 & 0 \\ -i\beta\gamma & 0\,0 & \gamma \end{pmatrix}; \tag{6.31}$$

the transposed matrix has β replaced by $-\beta$ and is clearly Λ^{-1}.

Next, note that the Λ matrix has the property that all its columns are orthogonal vectors, that is,

$$\sum_{\mu=1}^{4} \Lambda_{\mu\nu}\Lambda_{\mu\rho} = \delta_{\nu\rho}.$$

Hence,

$$\sum_{\mu} x'_\mu x'_\mu = \sum_{\mu,\nu} \Lambda_{\mu\nu}\Lambda_{\mu\chi} x_\nu x_\chi = \sum_{\nu} x_\nu x_\nu. \tag{6.32}$$

Since such sums occurs quite often, Einstein introduced the convention of understanding the summation symbol when the same index occur twice: so, one simply writes $s^2 = x_\nu x_\nu$. We shall adopt this convention.

[8] At first sight, it may seem odd that the components x_α measured in the system of the observer are differentiated with respect to the time measured in the reference of the moving body, but actually, $cd\tau = ds$ is an interval and does not depend on the reference system.

Let v_μ and w_μ be four-vectors; the scalar product is $v_\mu w_\mu$. A Lorentz transformation leads to

$$v'_\mu w'_\mu = \Lambda_{\mu\rho}\Lambda_{\mu\sigma} v_\rho w_\sigma, \tag{6.33}$$

but since $\Lambda_{\mu\rho}\Lambda_{\mu\sigma} = \delta_{\rho\sigma}$, we conclude that $v_\mu w_\mu$ is invariant, that is, it is a scalar. In particular, we have checked that $x_\mu x_\mu$ is the invariant square interval. From the four-velocity one obtains the invariant

$$v_\alpha v_\alpha = \gamma^2(u^2 - c^2) = -c^2. \tag{6.34}$$

Differentiating again, we find $v_\alpha \frac{dv_\alpha}{d\tau} = 0$, so the four-acceleration is orthogonal to the four-velocity.

Other quantities of pre-relativistic Physics are recognized as scalars or four-vector components in Relativity. Consider an electromagnetic wave. The wavelength is given by $\lambda = cT$, where T is the period; $\omega = ck$, where $k = \frac{2\pi}{\lambda}$ is the absolute value of the wave vector $\overrightarrow{k}$, pointing along the propagation direction $vers(\overrightarrow{E} \wedge \overrightarrow{H})$. Consider any component of the electromagnetic field of a wave in vacuo; its amplitude $f(\omega t - \overrightarrow{k} \cdot \overrightarrow{x})$ satisfies the wave equation $\frac{1}{c^2}\frac{\partial^2}{\partial t^2} f = \nabla^2 f$. In principle, an observer in any inertial reference can measure the phase difference between two space–time points by simply counting the maxima or the zeros of f between them, and counting is an invariant operation. The phase difference is a relativistic invariant, because it is just a matter of counting. So, $\Phi = k_\mu x_\mu$ is invariant, and

$$k_\mu = \left(\overrightarrow{k}, k_0 = \frac{\omega}{c}\right) \tag{6.35}$$

is a four-vector (called the wave four-vector). Among the other relevant four vectors, there is the current $j_\mu = (\overrightarrow{J}, i\frac{\rho}{c})$, where ρ is the charge density, and the four-potential

$$A_\mu = (\overrightarrow{A}, i\frac{\phi}{c}).$$

6.3.1 The Lorentz Transformation as a Rotation

There are other benefits from the formal analogy between (6.28) and a rotation in the plane, that is (leaving aside the y and z axes), in ordinary $\mathbb{R}^2$, where the rotation matrix is

$$\begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}.$$

Equation (6.28) corresponds to a rotation if we set $\cos(\theta) = \gamma$, $\sin(\theta) = i\beta\gamma \Rightarrow \tan(\theta) = i\beta$. Setting $\theta = i\xi$, since $\tan(i\xi) = i\tanh(\xi)$, we are left with $\tanh(\xi) =$

β. The special Lorentz transformation is a rotation by $\theta = i\xi$ in the $x_1 - x_4$ plane by an imaginary angle $i\xi$; this is such that $\tanh(\xi) = \beta$. The rotation is such that

$$x_1' = x_1 \cosh(\xi) + i x_4 \sinh(\xi) \quad x_4' = x_4 \cosh(\xi) - i x_1 \sinh(\xi).$$

The analogy with a rotation in an imaginary angle in $\mathbb{R}^4$ is not tremendously enlightening about the physical meaning, but it allows us to combine real rotations (which are also represented by 4×4 matrices) with special Lorentz transformations to make general Lorentz transformations in any directions,[9] which is the set of all possible Lorentz transformations and the rules to combine two of them. Moreover, this analogy helps us to find and classify the various kinds of invariant, which can be built in analogy for rotations and Lorentz transformations. These are the tensors.

6.3.2 Vectors and Tensors

When we work in $\mathbb{R}^3$ we must ensure that the physical equations remain valid if one rotates or reflects the Cartesian reference. This occurs trivially with scalars since their only component never changes; pseudo-scalars change sign under reflection. However, there are also vectors, like the electric field $\boldsymbol{E}$, and pseudo-vectors like the magnetic field $\boldsymbol{B}$ (pseudo because it does not change under parity). Now, if one rotates the reference, the components change in such a way that $\boldsymbol{E}$ and $\boldsymbol{B}$ remain the same, as they should. Two-index tensors are arrays with 2×3 components that transform under specific rules and in general any physical, measurable object must be represented by a tensor with some number of indices.

Chronotopic Tensors are similar to those in $\mathbb{R}^3$, except that in Relativity one must work in $\mathbb{R}^4$; they represent physical quantities that are the same for all observers. So, our equations must hold independently of rotations and boosts (i.e., Lorentz transformations).

> Einstein mathematically formulated the Principle of Relativity as follows:
> *The laws of Physics are equations between tensors.*

Thus we also gain a simple criterion to *relativize* the physical laws: they must be rewritten in covariant form, i.e., as equalities among tensors. One obtains laws that are good candidates since they (1) satisfy the principle of Relativity and (2) have the correct classical limits. In many cases this strategy is rewarding.

A relativistic invariant like the charge or the rest mass of a particle is the simplest type of tensor. It is a single quantity $\phi(x)$ (that may depend on the point x_μ) but is the same for all the inertial observers. It is also called a scalar or a zero-rank tensor (no indices). A four-vector $v_\mu(x)$, $\mu = 0, \ldots, 4$ Lorentz-transforms like x_μ and is a tensor with rank 1 and one index. Out of two four-vectors $v_\mu(x)$, $w_\nu(x)$ one can form

[9] The infinite set of matrices so obtained are called a representation of the Lorentz Group.

an invariant, like $\phi(x) = v_\mu(x)w_\mu(x)$ which is their scalar product in analogy with scalar products in $\mathbb{R}^3$. We have just seen that $x_\mu x_\mu$ is invariant.

Now consider the gradient of a scalar ϕ. It is $d\phi = \frac{\partial\phi}{dx_\mu}dx_\mu$. Since $\phi(x)$ is a scalar and dx_μ are the components of a vector (actually, for short, one says that dx_μ is a vector), it is clear that the gradient $\frac{\partial\phi}{dx_\mu}$ is a four-vector obtained by differentiation. A tensor of rank 2 $w_{\mu\nu}$, $\quad(\mu = 1, \ldots, 4, \nu = 1, \ldots, 4)$, transforms like $x_\mu x_\nu$, that is, like the products of two four-vectors:

$$w'_{\mu\nu} = \Lambda_{\mu\rho}\Lambda_{\nu\sigma}w_{\rho\sigma}.$$

Higher rank tensors (i.e., ones with three or more indices) are defined similarly. It is fairly obvious that the sum of two tensors of the same rank is another tensor and so one can take linear combinations. A tensor T_{ik} such that $T_{ik} = T_{ki}$ is called a symmetric tensor; if instead $T_{ik} = -T_{ki}$ the tensor is antisymmetric[10]; the importance of such properties is evident if one considers that they are preserved by Lorentz transformations. So, for example, an antisymmetric tensor is such for all observers. By doing the derivatives, one can easily check that the electromagnetic field is an antisymmetric tensor:

$$F_{\mu,\nu} = \frac{\partial A_\nu}{\partial x_\mu} - \frac{\partial A_\mu}{\partial x_\nu}. \tag{6.36}$$

This can be written as an antisymmetric matrix:

$$F = \begin{pmatrix} 0 & B_z & -B_y & -iE_x \\ -B_z & 0 & B_x & -iE_y \\ B_y & -B_x & 0 & -iE_z \\ iE_x & iE_y & iE_z & 0 \end{pmatrix}. \tag{6.37}$$

Thus, we know how to transform the electromagnetic fields. For the special transformation (6.30), one finds

$$\begin{aligned} B'_x &= B_x & \qquad E'_x &= E_x \\ B'_y &= \gamma(B_y + \beta E_z) & \qquad E'_y &= \gamma(E_y - \beta B_z) \\ B'_z &= \gamma(B_z - \beta E_y) & \qquad E'_z &= \gamma(E_z + \beta H_y). \end{aligned} \tag{6.38}$$

This is the fast way to transform the field: otherwise, one can Lorentz-transform the charge-current four-vector and then re-calculate the fields using Maxwell's equations. It is easy to realize that if in the unprimed system, a charge is acted upon by an electric field, in the primed system a Lorentz force also arises. From the field tensor, one can form a third rank tensor

[10] In general, one can form tensors that belong to irreducible representations of the Group of permutations of the indices; such a property is invariant.

$$\frac{\partial}{\partial x_\lambda} F_{\mu\nu} = \frac{\partial^2 A_\nu}{\partial x_\lambda \partial x_\mu} - \frac{\partial^2 A_\mu}{\partial x_\lambda \partial x_\nu}. \tag{6.39}$$

Two more tensors can be obtained by permutation of indices:

$$\frac{\partial}{\partial x_\mu} F_{\nu\lambda} = \frac{\partial^2 A_\lambda}{\partial x_\mu \partial x_\nu} - \frac{\partial^2 A_\nu}{\partial x_\mu \partial x_\lambda} \tag{6.40}$$

and

$$\frac{\partial}{\partial x_\nu} F_{\lambda\mu} = \frac{\partial^2 A_\mu}{\partial x_\nu \partial x_\lambda} - \frac{\partial^2 A_\lambda}{\partial x_\nu \partial x_\mu}. \tag{6.41}$$

These are not independent, because their sum reads as

$$\frac{\partial}{\partial x_\lambda} F_{\mu\nu} + \frac{\partial}{\partial x_\mu} F_{\nu\lambda} + \frac{\partial}{\partial x_\nu} F_{\lambda\mu} = 0. \tag{6.42}$$

To see the meaning, let us introduce the fields: what we got is actually $div\boldsymbol{B} = 0$. Moreover, $\frac{\partial F_{12}}{\partial x_4} + \frac{\partial F_{24}}{\partial x_1} + \frac{\partial F_{41}}{\partial x_2} = 0$ is nothing but $\frac{1}{c}\frac{\partial}{\partial t} B_3 + \frac{\partial E_2}{\partial x_1} - \frac{\partial E_1}{\partial x_2} = 0$, which is a component of the $\nabla \wedge \boldsymbol{E}$ Maxwell equation; the other components are obtained by permutation of the indices. The inhomogeneous Maxwell equations are rewritten as

$$\frac{\partial F_{\mu\nu}}{\partial x_\nu} = \frac{4\pi}{c} J_\mu. \tag{6.43}$$

We are in a position to Lorentz-transform the fields. Let $\boldsymbol{E}_\parallel$, $\boldsymbol{B}_\parallel$ denote the field components parallel to the speed $\boldsymbol{v}$ and $\boldsymbol{E}_\perp$, $\boldsymbol{B}_\perp$ the orthogonal components.

Problem 6.5 The relativistic law of sum of speeds must result from the sum of four-velocities. Verify that.

Solution 6.5 If in the train reference $\mathcal{K}_\mathcal{T}$, a material point moves with three-velocity $\overrightarrow{v}^T = (v, 0, 0)$ the train driver finds that the four-velocity is

$$w_\mu^{(T)} = \frac{1}{\sqrt{1 - \frac{v^{T2}}{c^2}}} (\overrightarrow{v}^T, ic). \tag{6.44}$$

Lorentz-transforming, we find the four-vector in $\mathcal{K}_\mathcal{S}$; since the speed of the train is u,

$$w_1^{(S)} = \frac{w_1^{(T)} - \frac{iu}{c} w_4^{(T)}}{\sqrt{1 - \frac{u^2}{c^2}}},$$

and substituting,

$$w_1^{(S)} = \frac{1}{\sqrt{1 - \frac{v^{T2}}{c^2}}} \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}} (v^T + u).$$

For coherence with (6.44) $w_1^{(S)} = \frac{v^S}{\sqrt{1-\frac{v^{S2}}{c^2}}}$, where v^S is measured in the station. So,

$$\frac{(v^S)^2}{1-\left(\frac{v^S}{c}\right)^2} = \varphi, \qquad \varphi = \frac{(u+v)^2}{\left(1-\left(\frac{u}{c}\right)^2\right)\left(1-\left(\frac{v}{c}\right)^2\right)},$$

which implies

$$(v^S)^2 = \frac{\varphi}{1+\frac{\varphi}{c^2}} = \frac{(u+v)^2}{\left(1-\left(\frac{u}{c}\right)^2\right)\left(1-\left(\frac{v}{c}\right)^2\right)+\frac{(u+v)^2}{c^2}}.$$

The denominator is

$$1-\left(\frac{u}{c}\right)^2-\left(\frac{v}{c}\right)^2+\left(\frac{uv}{c^2}\right)^2+\left(\frac{u}{c}\right)^2+\left(\frac{v}{c}\right)^2+2\frac{uv}{c^2} = \left(1+\frac{uv}{c^2}\right)^2.$$

Finally, one finds the known result:

$$v^S = \frac{v^T+u}{1+\frac{v^T u}{c^2}}.$$

Problem 6.6 A spaceship travels with speed v relative to a space station. The pilot puts a mirror A ahead and a mirror B on the tail at a distance $AB = 2l_0$; a light source S in the middle emits a signal that hits A and B simultaneously after a time τ_0 measured on the clock of the spaceship. An observer in the space station finds that the signal reaches B first and then A after a time τ_0. Find the speed of the space ship.

Solution 6.6 In the station, the space ship is long $l = l_0\gamma$; the time to reach A is $\tau_A = \frac{l}{c-v}$, while the time to reach B is $\tau_A = \frac{l}{c+v}$; so $\Delta t = 2\frac{l_0}{c}\frac{u}{\sqrt{c^2-u^2}}$ must be equated to $\frac{l_0}{c}$. So one finds that $\frac{v}{c} = \frac{1}{\sqrt{5}}$.

6.3.3 *Action of the Free Field*

The Lagrangian of the field can be taken to be proportional to $F_{\mu\nu}^2$; it is usually written in the form

$$L = \frac{i}{4\pi c}\left(\frac{\partial A_\nu}{\partial x_\mu}-\frac{\partial A_\mu}{\partial x_\nu}\right)\left(\frac{\partial A_\nu}{\partial x_\mu}-\frac{\partial A_\mu}{\partial x_\nu}\right). \tag{6.45}$$

Of course, the factor $\frac{i}{4\pi c}$ is arbitrary. The Lagrange equations are

$$\frac{\partial L}{\partial A_\mu} = \frac{\partial}{\partial x_\nu}\frac{\partial L}{\partial\left(\frac{\partial A_\mu}{\partial x_\nu}\right)}.$$

One finds that

$$\frac{\partial}{\partial x_\nu}\left(\frac{\partial A_\nu}{\partial x_\mu}-\frac{\partial A_\mu}{\partial x_\nu}\right)=\frac{\partial}{\partial x_\nu}F_{\mu\nu}=0\Leftrightarrow divB=0, rotE+\frac{1}{c}\frac{\partial B}{\partial t}=0.$$

6.3.4 Doppler Effect and Aberration of Light

The Doppler effect is a phenomenon that occurs when there is a relative motion between a source of waves and an observer. It causes a shift in the frequency and wavelength of the waves, which can be observed as a change in pitch or color.

It was first described by the Austrian physicist Christian Doppler in 1842. Doppler noticed that the pitch of the sound produced by a train horn seemed to change as the train passed by, becoming higher in frequency as it approached and lower in frequency as it moved away, and gave an explanation of the effect. Here, we are interested in the relativistic theory of the optical Doppler effect, which corrects the previous explanation in important details.

Let us conduct a new thought experiment, with the help of the now usual train running along the x axis with velocity u. The station is replaced by a monochromatic light source that emits a plane electromagnetic wave with pulsation $\omega_S = 2\pi\nu_S$ and wave vector $\overrightarrow{k}_S$. The index S stands for source. The modulus of $\overrightarrow{k}_S$ is $k_S = \frac{\omega_S}{c}$ and the direction makes an angle θ_S with the x axis, therefore the component along $\overrightarrow{u}$ is

$$(k_S)_\parallel = (\overrightarrow{k}_S)_x = k_S\cos\theta_S = \frac{\omega_S}{c}\cos\theta_S,$$

while the vertical component is

$$(k_S)_\perp = k_S\sin\theta_S = \frac{\omega_S}{c}\sin\theta_S. \tag{6.46}$$

In the reference of the train, the driver observes an electromagnetic wave, but we must find out which wave it is. Let us say, it has some angular frequency ω_T; its wave vector has some module $k_T = \frac{\omega_T}{c}$, and it makes some angle θ_T with the x axis, and therefore its components are

$$(k_T)_\parallel = (\overrightarrow{k}_T)_x = \frac{\omega_T}{c}\cos\theta_T,\ (k_T)_\perp = \frac{\omega_T}{c}\sin\theta_T. \tag{6.47}$$

In order to set up the relativistic theory, we must recognize the scalars and tensors involved in the propagation of light.

One can start by noting that the phase $\Phi = \boldsymbol{k}.\boldsymbol{r} - \omega t$ must be invariant, since the number of nodes of the field between two marks is something objective and cannot depend on the reference. We have seen that there are no transformation rules for three-component $\overrightarrow{k}$ vectors in Relativity, but we know how to deal with transforming

scalars, four-vectors and other tensors. Therefore a fourth component is wanted. It is easily understood that the four-vector must be

$$k_\mu = \left(\overrightarrow{k}\,, k_0 = \frac{\omega}{c}\right) \tag{6.48}$$

such that we obtain the phase as the scalar product $k_\mu x_\mu = \Phi$.

In the reference of the train, the driver observes an electromagnetic wave, but we must find out which wave it is. Let us say, it has some angular frequency ω_T; its wave vector has some module $k_T = \frac{\omega_T}{c}$, and it makes some angle θ_T with the x axis, and therefore its components are

$$(k_T)_\parallel = (\overrightarrow{k}_T)_x = \frac{\omega_T}{c}\cos\theta_T,\ (k_T)_\perp = \frac{\omega_T}{c}\sin\theta_T.$$

The train driver draws a line, makes two marks on it and—at least in principle—he can count the nodes (points with vanishing field) of the wave between marks at a given instant. The important remark is that the number of nodes between two marks cannot depend on the reference.

Therefore, the phase $\Phi = \boldsymbol{k}.\boldsymbol{r} - \omega t$ must be invariant. Since x_μ is a four-vector, and is involved, the phase must be the scalar product with another four-vector, which generalizes the pre-relativistic wave vector. This is evidently k_μ. Therefore,

$$k_\mu = \left(\overrightarrow{k}\,, k_0 = \frac{\omega}{c}\right).$$

The Lorentz transformation written in real form (6.11) is

$$\begin{cases} (k_T)_\parallel = \gamma\left[(k_S)_\parallel - \beta(k_S)_0\right] & \Longrightarrow \omega_T\cos\theta_T = \gamma\omega_S\left[\cos\theta_S - \beta\right], \\ (k_T)_\perp = (k_S)_\perp & \Longrightarrow \omega_T\sin\theta_T = \omega_S\sin\theta_S, \\ (\overrightarrow{k}_T)_0 = \gamma((k_S)_0 - \beta(k_S)_\parallel) & \Longrightarrow \omega_T = \gamma\omega_S\left[1 - \beta\cos\theta_S\right]. \end{cases} \tag{6.49}$$

The last line of Eq. (6.49) gives the frequency on the train in terms of the the frequency and the angle in the station, while the first two lines yield

$$\tan(\theta_T) = \frac{\sin(\theta_S)}{\gamma(\cos(\theta_S) - \beta)}. \tag{6.50}$$

By plugging $\cos(\theta_S)$ derived from the first and simplifying, after some calculations one obtains the formula found by Einstein in 1905:

$$\omega_T = \frac{\omega_S}{\gamma[1 + \beta\cos(\theta_T)]}. \tag{6.51}$$

6.3.4.1 The Aberration of Light

The source is seen in a direction that depends on the speed of the observer relative to the source. This phenomenon is known as the aberration of the light and has been well known to astronomers since the eighteenth century. It was discovered by the Englishman James Bradley (1693–1762), later the Royal Astronomer. It is the apparent motion of all the stars due to the revolution of the Earth around the Sun. From the first line of Eq. (6.49), one obtains $\cos\theta_T = \frac{\gamma\omega_S}{\omega_T}(\cos(\theta) - \beta)$; plugging $\frac{\gamma\omega_S}{\omega_T}$ from the third line, one arrives at

$$\cos(\theta_T) = \frac{\cos(\theta_S) - \beta}{1 - \beta\cos(\theta_S)}. \tag{6.52}$$

In this case the train is our Earth; we see the stars shifted toward the direction of the motion of the Earth due to the orbital motion.

The effect changes the apparent position of stars by 20 arc-s when the line of sight is perpendicular to the Earth's orbit, and is a direct evidence that our Planet rotates around the Sun. Of course there is also a secular shift due to the motion of the sun in the Galaxy. A consequence is that a fast forward-moving observer sees objects grouped in the forward direction rather than in a rearward one; this is referred to as the "searchlight" or "headlight" effect.

6.3.4.2 The Longitudinal Doppler Effect

Returning to Eq. (6.49), for $\theta_S = 0$, the observer moves away from the source and experiences the longitudinal Doppler effect[11]:

$$\omega_T = \omega_S\sqrt{\frac{1 - \frac{u}{c}}{1 + \frac{u}{c}}}. \tag{6.53}$$

On the other hand, if the train goes toward the light source, $\theta_S = \pi$, there is a blue shift which is a first-order effect for small speeds. Qualitatively, this is similar to the behavior of sound waves that reach us from moving objects. This effect allows for precise measurement of speeds and is of paramount importance in Astrophysics and Cosmology. Edwin Hubble, in 1929, discovered that the light from all the far Galaxies reddens with increasing distance and concluded that the Universe is expanding (see Sect. 8.12). The Doppler effect also allows the police to measure the speed of cars and fine the drivers who exceed the limits.

[11] Here u is the speed of the train; more usually the law is written in terms of the speed $-u$ of the source relative to the observer.

6.3.4.3 Black Body

A moving observer sees a black body as a black body, even though it will be hotter for an approaching observer and colder for one moving away from the body. This is an obvious consequence of the Planck formula, which depends on $\frac{h\nu}{K_B T}$, but one of paramount physical importance.

6.3.4.4 The Transverse Doppler Effect and the Mösbauer effect

Equation (6.49) also shows that for $\cos(\theta_S) = \beta$, $\theta_T = \frac{\pi}{2}$, and

$$\omega_T = \gamma\omega_S \left[1 - \beta^2\right] = \omega_S\sqrt{1 - \beta^2}.$$

This *transverse Doppler effect* is second-order in β, and is not predicted by the pre-relativistic Physics. It can be understood as being due to the relativistic time dilation.

Accurate experimental verifications have been done using the recoilless emission discovered by Rudolf Mössbauer. The Mössbauer effect is a solid state effect. Typically, a free ^{57}Fe atom makes a nuclear decay emitting a γ ray and the transition is sharp, but the γ is broadened because part of the energy goes to the recoil of the nucleus in some direction. Mössbauer discovered that if the atom is bound within a crystal, it emits a very narrow γ ray. The vibrations in solids are quantized harmonic oscillators, called phonons. Since the oscillator state with the recoiling atom is not orthogonal to the crystal ground state, there is a probability that a nuclear decay emits a γ ray without exciting the oscillator at all. In this case, the whole crystal, rather than the Fe ion, recoils. Since the crystal is macroscopic, in practice a recoilless emission occurs. This effect allows to work with very well-defined frequencies, since it avoids the recoil that is an important source of broadening. Doppler shifts due to velocities of the order of 1 cm/s can be measured.[12] The relativistic treatment of the Doppler effect, combined with the quantum theory of the black body radiation to be developed below, has another very interesting consequence. If a hot body emits a black body spectrum at some temperature T, an observer moving toward it will see a spectrum shifted to the blue while an observer moving away from it will see a spectrum shifted to the red, but both will still observe a black body spectrum; only the temperatures change. The fact that a body is black cannot depend on the velocity of the observer.

[12] For the experimental confirmation of the transverse Doppler effect, see Hay et al. (1960).

6.4 Relativistic Mechanics

The non-relativistic mechanics is very good in everyday life and also for sending a rocket to Pluto, but not for particles in the Large Hadron Collider.[13] The Lagrangian formulation must be extended to the relativistic case. Let an inertial observer describe the simplest possible system, that is, the motion of a free particle in the interval of time from t_1 to t_2. We must have

$$\delta S = \delta \int_{t_1}^{t_2} L(q(t), \dot{q}(t))dt = 0, \tag{6.54}$$

where L is a relativistic Lagrangian to be found, and the time is the one measured in the reference of the observer. Since the particle's proper time $\frac{ds}{c} = d\tau = \sqrt{1 - \left(\frac{v}{c}\right)^2}dt$ is scalar, it is natural to take

$$S = -m_0 c^2 \int_{t_a}^{t_b} \sqrt{1 - \left(\frac{v}{c}\right)^2}\, dt; \tag{6.55}$$

the constant $-m_0c^2$ is a good choice, since for $v \ll c$, L gives back the non-relativistic kinetic energy: $L \approx m_0c^2 + \frac{1}{2}m_0v^2 + \cdots$. The constant m_0c^2 does not change the equations of motion and can be ignored. The momentum conjugated with $\overrightarrow{x}$ is

$$\overrightarrow{p} = \frac{\partial L}{\partial \overrightarrow{v}} = \frac{m_0 \overrightarrow{v}}{\sqrt{1 - \left(\frac{v}{c}\right)^2}} \equiv m \overrightarrow{v}, \tag{6.56}$$

with m the relativistic mass. This reduces[14] to the Galilean momentum for small v. The change is that the mass m is velocity-dependent, $m \approx m_0$ for $v \to 0$, but m diverges for $v \to c$, So, c is a limiting velocity, unattainable for massive bodies. Nobody can overtake a light ray and look back to see what it looks like! The energy is given by the Lagrangian formula $E = \overrightarrow{p} \cdot \overrightarrow{v} - L$. One finds

$$E = \frac{m_0c^2}{\sqrt{1 - \left(\frac{v}{c}\right)^2}} \equiv mc^2. \tag{6.57}$$

[13] Theoretical Physics advances by successive generalizations. The new formulations do not contradict the older ones, that remain valid as special cases, but extend them to new situations. This is something that some people, including philosophers, cannot understand. Science does not have paradigms and does not follow fashion, as some epistemologists say. Eventually they criticize Science because of its most important feature: the capacity to make progress.

[14] Using $\frac{1}{\sqrt{1-x^2}} \approx 1 + \frac{x^2}{2} + \frac{3x^4}{8}$.

This formula is so famous that even the man of the street knows it. It says that a fixed mass m has a *rest energy*

$$\boxed{E_0 = m_0 c^2,}$$

while a moving mass also has kinetic energy; at low speed, this is

$$\boxed{E \approx m_0 c^2 + \frac{1}{2} m_0 v^2 + \cdots}$$

which coincides with the Galilean expression plus the rest energy. Like the momentum, the energy E blows up at c. But all that is really striking only when it is properly understood, because in Newtonian Mechanics, the origin of energy is arbitrary and only energy differences are meaningful. The real point is that in Relativity theory, there is no such arbitrariness. Indeed, the momentum

$$p_\mu = m_0 w_\mu = \frac{m_0}{\sqrt{1 - \left(\frac{v}{c}\right)^2}} (v, ic) = \left(\vec{p}\,, i\frac{E}{c}\right) \tag{6.58}$$

is a four-vector, and this property would be broken if one could add a constant to the fourth component. This implies that we can know the energy content of a system from its rest mass, and conversely, any amount of energy has inertia. That was a big change in Science. An obvious consequence is that the mass of bound systems (like nuclei) is less than the sum of the masses of the constituents, since in the nucleo-synthesis, some energy is emitted as γ rays.

By writing the energy (6.57) in terms of the four-momentum, we obtain the Hamiltonian

$$H = c\sqrt{p^2 + m^2 c^2}. \tag{6.59}$$

The Hamiltonian formalism is not altered by the process of relativization. The Hamilton–Jacobi equation has the relativistic counterpart

$$\frac{1}{c^2}\left(\frac{\partial S}{\partial t}\right)^2 - \left(\frac{\partial S}{\partial x}\right)^2 - \left(\frac{\partial S}{\partial y}\right)^2 - \left(\frac{\partial S}{\partial z}\right)^2 = m^2 c^2. \tag{6.60}$$

Thus the relativistic formalism is, *mutatis mutandis*, much like the usual one.

These findings are readily extended to deal with a charged particle in a field. The pre-relativistic Lagrangian is shown in (2.88). Once we modify the kinetic energy term as shown above, we get

$$L(\vec{r}\,, \vec{v}\,, t) = -m_0 c^2 \sqrt{1 - \frac{v^2}{c^2}} - q\phi + q\,\vec{v} \cdot \vec{A}\,(\vec{r}\,, t). \tag{6.61}$$

Now one can derive the relativistic equations of motion of a charge in a field

$$\frac{d}{dt}\overrightarrow{p} = \overrightarrow{F}, \quad \overrightarrow{p} = m_0\gamma\overrightarrow{v}, \tag{6.62}$$

where $\overrightarrow{F}$ is the Lorentz force (2.85).

Equation (6.62) must be put into a manifestly covariant form. The derivative must be $\frac{d}{d\tau}$, with respect to the proper time τ, and the force must become a four-vector. The equations of Dynamics become

$$\frac{d}{d\tau}p_\mu = f_\mu. \tag{6.63}$$

Multiplying by $v_\mu = \frac{1}{\sqrt{1-(\frac{v}{c})^2}}(\overrightarrow{v}, ic)$ and recalling that $v_\mu v_\mu = -c^2$, one finds

$$m_0 v_\mu \frac{d}{d\tau}v_\mu = \frac{m_0}{2}\frac{d}{d\tau}(v_\mu v_\mu) = 0 = f_\mu v_\mu.$$

So,

$$\overrightarrow{f}\cdot\overrightarrow{v} + if_4 c = 0,$$

$$f_4 = \frac{i\overrightarrow{f}\cdot\overrightarrow{v}}{c}.$$

Since $dt = \gamma d\tau$ and $\frac{d}{d\tau} = \gamma\frac{d}{dt}$, the space part of (6.63) becomes

$$\frac{d}{d\tau}\overrightarrow{p} = \frac{1}{\sqrt{1-\left(\frac{v}{c}\right)^2}}\frac{d}{dt}\overrightarrow{p} = \overrightarrow{f} = \frac{1}{\sqrt{1-\left(\frac{v}{c}\right)^2}}\overrightarrow{F},$$

and this relates the modified three-vector $\overrightarrow{f}$ with the Lorentz force $\overrightarrow{F}$ (2.85).

Moreover,

$$f_4 = \frac{i}{c}\frac{\overrightarrow{F}\cdot\overrightarrow{v}}{\sqrt{1-\left(\frac{v}{c}\right)^2}},$$

and the four-force is $f_\mu = \left(\frac{\overrightarrow{F}}{\sqrt{1-(\frac{v}{c})^2}}, \frac{i}{c}\frac{\overrightarrow{F}\cdot\overrightarrow{v}}{\sqrt{1-(\frac{v}{c})^2}}\right)$. The fourth component of the equations of motion is

$$\frac{d}{dt}\frac{m_0c^2}{\sqrt{1-\left(\frac{v}{c}\right)^2}} = \overrightarrow{F}\cdot\overrightarrow{v}.$$

In this way, the equations of Dynamics are relativized.

6.4.1 Energy–Momentum Tensor (Alias Stress–Energy Tensor)

In the chapter on General Relativity, we will need the Energy–Momentum Tensor $T^{\mu\nu}$ (also known as the Stress–Energy Tensor) that represents the energy density, pressure and other properties of a fluid comprising particles and fields. This is a symmetric tensor ($T^{\mu\nu} = T^{\nu\mu}$) that, broadly speaking, represents the flux of p^μ across a surface of constant x^ν plus the flux of p^ν across a surface of constant x^μ.

Consider a fluid having proper density (that is, measured in the rest reference) $\rho_0(x)$ in the chronotopic point x, and four-velocity u_μ (also a function of space–time) with negligible pressure. We may take the *stress–energy tensor* in the form

$$T^{\mu\nu} = \rho_0(x) u^\mu u^\nu. \tag{6.64}$$

The divergence may be denoted by $T^{\mu\nu}_{\ ,\mu} \equiv \frac{\partial T^{\mu\nu}}{\partial x^\mu} \equiv \partial_\mu T^{\mu\nu}$. At any rate,

$$T^{\mu\nu}_{\ ,\mu} = \partial_\mu(\rho_0 u^\mu) u^\nu + \rho_0 u^\mu \partial_\mu u^\nu.$$

This can be simplified, since in the rest frame the velocity is zero and the second term vanishes; moreover, $\partial_\mu(\rho_0 u^\mu) = 0$ by the continuity equation. Thus,

$$\boxed{T^{\mu\nu}_{\ ,\mu} = 0,} \tag{6.65}$$

in other words, the energy–momentum tensor is *divergenceless*. This equation expresses the conservation of energy and momentum in Relativity and will turn out to be important for General Relativity.

If the fluid has proper internal energy U_0 and proper pressure P_0, extra contributions arise to the stress–energy tensor. One can show that a more complete, still divergenceless expression, is

$$T_{\mu\nu} = \delta_{\mu\nu} P_0 + \left(\rho_0(x) + \frac{\rho_0 U_0}{c^2} + \frac{P_0}{c^2}\right) u_\mu u_\nu. \tag{6.66}$$

This is a covariant way to represent the energy–matter contents of space–time and is crucially important for General Relativity.

References

Hay, et al.: Phys. Rev. Lett. **4**, 165 (1960)
Lampa, A.: Zeitschrift für Phys. **27**, 138–148 (1924)

Chapter 7
Curvilinear Coordinates and Curved Spaces

To proceed, we must develop some mathematical tools drawn from Differential Geometry.

Even in flat Euclidean space it may be useful to use curvilinear coordinates; for instance, in three-dimensional problems having central symmetry, we obtain an important simplification when the line element $ds^2 = dx^2 + dy^2 + dz^2$ is replaced by $ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2(\theta) d\phi^2$. In a curved space we have no other choice, because Cartesian coordinates may exist only locally, that is, in an infinitesimal neighborhood. One example is the surface of a sphere of radius R. The spherical coordinates with r set equal to the radius R of the sphere do the job. In the latter case, we deal with a two-dimensional curved subspace embedded in a three-dimensional Euclidean space. The curved coordinates are intrinsic to the surface, and one can ignore the existence of a radial dimension.

The Minkowsky space is not general enough for our purposes. We need a Riemannian space, which has a curvature. Since we live on the curved surface of a more or less spherical planet, such ideas are not too far from common wisdom, but the mathematical apparatus that we need is far from trivial. The relevant theory was created by outstanding mathematicians: Karl Friederich Gauss (1777–1855), Bernhard Riemann (1826–1866), Elvin Bruno Christoffel (1829–1900), Gregorio Ricci Curbastro (1853–1925) and Tullio Levi Civita (1873–1942).

It is advantageous, from a pedagogical point of view, to introduce curvilinear coordinates starting from curves in the plane, that is, in $\mathbb{R}^2$, since the extension to curved surfaces in $\mathbb{R}^d$ for any larger dimension d is then obtained easily. This is a good strategy: once we have developed the formalism, it fits our needs when working in curved spaces, where Cartesian coordinates do not exist, and in the actual space–time (Fig. 7.1).

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_7

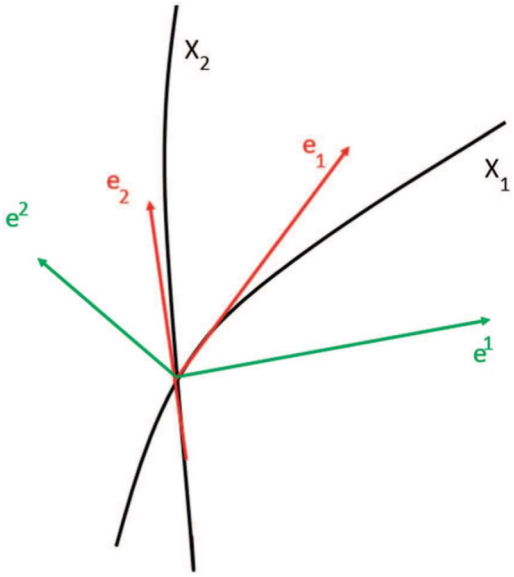


Fig. 7.1 When the coordinate lines are not orthogonal, the covariant basis vectors (red) are different from the contravariant ones (green) and do not have fixed length

In $\mathbb{R}^2$, we can choose an origin O and a Cartesian system such that any point may be assigned as $\overrightarrow{\boldsymbol{R}} = (X^1, X^2)$. The square distance from a given point $\overrightarrow{\boldsymbol{R}}$ to a near point

$$\overrightarrow{\boldsymbol{R}} + d\overrightarrow{\boldsymbol{R}} \tag{7.1}$$

is

$$ds^2 = dX_1^2 + dX_2^2. \tag{7.2}$$

Let $x^i = x^i(X^1, X^2)$ be curvilinear coordinates in terms of the Cartesian $\overrightarrow{\boldsymbol{R}} = (X^1, X^2)$. We assume that these functions are invertible, that is, $x^i = x^i(X^1, X^2) \Leftrightarrow X^i = X^i(x^1, x^2)$. If x^1 is varied while x^2 is kept fixed, $\overrightarrow{\boldsymbol{R}}$ will trace out a curve, the x^1 curve. In this way, one can think of a local coordinate system at each point with basis vectors

$$\overrightarrow{e}_i = \frac{\partial \overrightarrow{\boldsymbol{R}}}{\partial x^i} = \left(\frac{\partial X^1}{\partial x^i}, \frac{\partial X^2}{\partial x^i}\right). \tag{7.3}$$

These are called *covariant* basis vectors; note that the component i is a subscript. Now the infinitesimal shift ds^2 (7.2) becomes, understanding the sum over repeated indices,

$$ds^2 = \frac{\partial \boldsymbol{R}}{\partial x^i}\frac{\partial \boldsymbol{R}}{\partial x^j} dx^i dx^j = \boldsymbol{e}_i \boldsymbol{e}_j dx^i dx^j = g_{ij} dx^i dx^j,$$

where

$$g_{ik} = \overrightarrow{e}_i \overrightarrow{e}_k \tag{7.4}$$

is called the covariant component ik of the metric tensor. The Einstein convention now is that repeated indices are meant to be summed if they are a subscript and a superscript.

But there is the complication that, in general, the covariant basis vectors are not of unit length and not orthogonal; moreover, they do not coincide with the contravariant basis vectors which are defined by $\overrightarrow{e}^{\,i}.\overrightarrow{e}_{\,k} = \delta_{ik}$.

Let us choose a point and start from it along a curve $x^i = x^i(s)$; we can arrange that s is the curvilinear distance from the starting point: $ds^2 = d\overrightarrow{R}.d\overrightarrow{R}$. Then, $\overrightarrow{t} = \frac{\partial \overrightarrow{R}}{\partial s}$ is the velocity, that is, is tangent to the curve and $\overrightarrow{t}.\overrightarrow{t} = 1$. Note that

$$\overrightarrow{t} = \frac{\partial \overrightarrow{R}}{\partial s} = \frac{\partial \overrightarrow{R}}{\partial x^i}\frac{\partial x^i}{\partial s} = \overrightarrow{e}_{\,i}\frac{\partial x^i}{\partial s} \tag{7.5}$$

$$\overrightarrow{e}_{\,i}\overrightarrow{e}_{\,k}\frac{\partial x^i}{\partial s}\frac{\partial x^k}{\partial s} = 1 \Rightarrow ds^2 = \overrightarrow{e}_{\,i}\overrightarrow{e}_{\,k}dx^i dx^k. \tag{7.6}$$

As anticipated above, to proceed we need tensors. Using (7.3) the metric tensor components are

$$g_{ij} = \left(\frac{\partial X^1}{\partial x^i}, \frac{\partial X^2}{\partial x^i}, \frac{\partial X^3}{\partial x^i}\right).\left(\frac{\partial X^1}{\partial x^j}, \frac{\partial X^2}{\partial x^j}, \frac{\partial X^3}{\partial x^j}\right)$$

$$= \frac{\partial X^1}{\partial x^i}\frac{\partial X^1}{\partial x^j} + \frac{\partial X^2}{\partial x^i}\frac{\partial X^2}{\partial x^j} + \frac{\partial X^3}{\partial x^i}\frac{\partial X^3}{\partial x^j}.$$

So,

$$g_{ij} = \frac{\partial X^k}{\partial x^i}\frac{\partial X^k}{\partial x^j}. \tag{7.7}$$

From $g_{ii} = \overrightarrow{e}_{\,i}^{\,2}$, the length of $\overrightarrow{e}_{\,i}$ is

$$|\overrightarrow{e}_{\,i}| = \sqrt{g_{ii}}.$$

The angles between the basis vectors can be obtained as follows. From $g_{ik} = \overrightarrow{e}_{\,i}\overrightarrow{e}_{\,k} = |\overrightarrow{e}_{\,i}||\overrightarrow{e}_{\,k}|\cos(\theta_{ik})$, one finds

$$\cos(\theta_{ik}) = \frac{g_{ik}}{\sqrt{g_{ii}g_{kk}}}.$$

Any vector at the point O can be expanded: $\overrightarrow{A} = A^i\overrightarrow{e}_{\,i}$; in this case, A^i are the *contravariant* components and bring i as a suffix. Expand the contravariant vectors on the covariant basis

$$\overrightarrow{e}^{\,i} = g^{ik}\overrightarrow{e}_{\,k}, \tag{7.8}$$

where g^{ik} are the contravariant components of the metric tensor.

The theory uses scalars (i.e., functions of the point that do not depend on the coordinate system), vectors, that may be covariant or contravariant, and tensors that can have several high or contravariant indices and low or covariant ones. In Eq. (7.8) we succeeded to raise the covariant index k, which is a subscript, to a contravariant index i which is a superscript. To do so, we have multiplied by the contravariant component g^{ik} and contracted, i.e., summed over k. This turns out to be a general rule. The converse rule allows us to lower a contravariant index k by multiplying for g_{ik} and contracting over k; the index has become a covariant index i. The general rule for raising and lowering indices involves the metric tensor. Each index can be raised (converting from covariant to contravariant) or vice versa, and back. To convert the covariant components A_i of a vector, one raises the index by setting

$$A^i = g^{ij} A_j, \; A_i = g_{ij} A^j. \tag{7.9}$$

A similar rule allows to raise and lower indices in tensors with any number of *legs* (that is, indices): for instance, $A^i_{jk} = g_{js} A^{is}_k$. To find g^{ik}, multiply by $\overrightarrow{e}_l$:

$$\overrightarrow{e}_l \overrightarrow{e}^i = \delta^i_l = g^{ik} g_{kl}. \tag{7.10}$$

Thus, the contravariant components are the elements of the matrix inverse of the matrix of the covariant components:

$$\{g^{ij}\} = \{g_{ij}\}^{-1}. \tag{7.11}$$

Note that the matrix g^{ij} is the inverse of the g_{ij} matrix and consequently $g^i_j = \delta^i_j = \delta_{ij}$. As noted above, the components of the contravariant tensor are denoted by superscripts.

In d dimensions, we must be able to transform to new coordinates $x \equiv (x_1, \ldots, x_d)$ from old ones x'. The differentials transform according to

$$dx^i = \frac{\partial x^i}{\partial x'^j} dx'^j; \tag{7.12}$$

the Einstein convention (that repeated indices are meant to be summed over from 1 to d if they are a subscript and a superscript) is used. dx^i is a prototype contravariant vector, with the component labels, are written as superscripts; the general contravariant transformation rule is

$$A^i = \frac{\partial x^i}{\partial x'^j} A'^j. \tag{7.13}$$

On the other hand, consider a quantity $f(x^i)$ which is a scalar, i.e., does not change in the transformation of variables. The derivatives of a scalar transform according to the covariant rule

$$\frac{\partial f}{\partial x^i} = \frac{\partial f}{\partial x'^j}\frac{\partial x'^j}{\partial x^i}. \tag{7.14}$$

A set of d components A_i transforming this way are the covariant components of a vector. Note that a superscript in the denominator is counted as a subscript. A second-order contravariant tensor has d^2 components that transform according to

$$T^{mn} = \frac{\partial x^m}{\partial x'^i}\frac{\partial x^n}{\partial x'^j}T'^{ij}, \tag{7.15}$$

that is, like the product of two contravariant vectors. A second-order covariant tensor has d^2 components that transform like the product of two covariant vectors, $T_{mn} = \frac{\partial x'^i}{\partial x^m}\frac{\partial x'^j}{\partial x^n}T'_{ij}$. A mixed tensor with one superscript index and one subscript index transforms like the product of a covariant and a contravariant vectors, and so on. In general, tensors may have any number of indices.

To fix the ideas, I assumed at the beginning that we were introducing curvilinear coordinates on the plane; however, neither dimensionality nor the fact that we could have used Cartesian axes enters into the considerations made. Therefore, the results in this section hold generally.

7.1 Calculus in Curved Spaces

7.1.1 Parallel Transport

In general, the components of a vector in a curved space depend on the point for two reasons: one is that the vector may vary, and the other is that even a constant vector has components that vary because the coordinate axes depend on the point. A vertical vector in Rome does not remain vertical if parallel-transported in Washington. Let us study first the case of a constant vector W. The covariant basis vectors at different points are not parallel and differ in length; however, the change of $\overrightarrow{e}_i$ when shifting the point by an infinitesimal increment dx^j of the curvilinear coordinate x^j must be proportional to the shift. See Fig. 7.2, which presents a two-dimensional illustration of the parallel transport. The blue line represents a curve $x^i(s)$ and s is a parameter. The figure shows that when a vector W (green) at A is reproduced without change in B, a position shifted along the x_1 curve, its components undergo a change that is first-order in dx_1. The difference between the covariant basis vectors based at points that lie at infinitesimal distances is

$$d\overrightarrow{e}_i = \Gamma^k_{ij}dx^j\overrightarrow{e}_k, \tag{7.16}$$

with Γ^k_{ij} the Christoffel symbol of the second kind, also called affine connections. Γ^k_{ij} will be computed below.

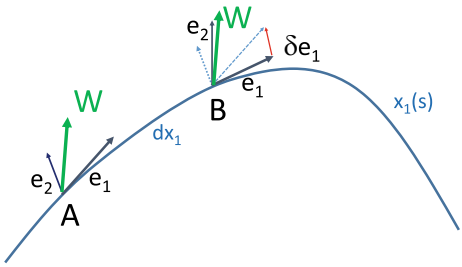


Fig. 7.2 Point A corresponds to the value x^1 of a coordinate line $x^1 = x^1(s)$ and is the origin of local basis vectors $\boldsymbol{e}_1 = \frac{dx^1}{ds}$ and $\boldsymbol{e}_2 = \frac{dx^2}{ds}$. Point B on the coordinate line corresponds to $x^i + dx^i$ and a new local basis. For a generic line, this differs from the dashed basis that would result from a simple shift of the basis at A, since the coordinate line is not straight. The change in $\boldsymbol{e}_1$ (shown in red) is of order dx

Consequently, the contravariant components of a constant vector W vary by

$$\delta W^i = -\Gamma^i_{kl} W^k dx^l. \tag{7.17}$$

Note that since $\frac{\partial \overrightarrow{e}_i}{\partial x^k} = \frac{\partial}{\partial x^k}\frac{\partial \overrightarrow{R}}{\partial x^i} = \frac{\partial}{\partial x^i}\frac{\partial \overrightarrow{R}}{\partial x^k} = \frac{\partial \overrightarrow{e}_k}{\partial x^i}$, it holds that

$$\Gamma^k_{ij} = \Gamma^k_{ji}.$$

Similarly, contravariant basis vectors at different points are not parallel, but the change of $\overrightarrow{e}^i$ when shifting the point by dx^j must be linear in the shift; when the change is expanded on the original basis, it is of the form

$$d\overrightarrow{e}^i = -\Gamma^i_{jk} dx^j \overrightarrow{e}^k; \tag{7.18}$$

the fact that the coefficients are indeed $-\Gamma^i_{jk}$ may be verified starting from $\overrightarrow{e}^i . \overrightarrow{e}_k = \delta_{ik}$ which implies that $\frac{\partial}{\partial x_j}(\overrightarrow{e}^i . \overrightarrow{e}_k) = 0 = (\frac{\partial \overrightarrow{e}^i}{\partial x^j}) . \overrightarrow{e}_k + \overrightarrow{e}^i . (\frac{\partial \overrightarrow{e}_k}{\partial x^j})$.

7.1.2 Curvature

The covariant components of a constant vector W vary by

$$\delta W_i = \Gamma^k_{il} W_k dx^l. \tag{7.19}$$

This implies that the parallel transport on a closed circuit produces a cumulative change of a constant vector by

$$\Delta W_i = \oint \Gamma^k_{il} W_k dx^l. \tag{7.20}$$

Imagine a pendulum swinging at the North Pole along the meridian of Greenwich. If it is carefully taken to the Equator by parallel transport, it will keep swinging along the meridian of Greenwich; now, if it is taken along the Equator to the meridian of Rome and then taken back to the North Pole, it will be found to oscillate along the meridian of Rome. This is clearly a consequence of the curvature of the Earth. By such an experiment, a two-dimensional being living on the surface of a sphere could discover its curvature and even measure the radius. We shall soon find that actually Eq. (7.33) is the starting point for defining the curvature tensor.

7.1.3 Christoffel Symbols

It easy to generalize the above discussion about parallel transport to variable vectors.

The del operator

$$\nabla = \overrightarrow{e}^i \frac{\partial}{\partial x^i} \tag{7.21}$$

is an invariant, since by contracting vectors, one obtains scalars. This feature is useful when one wishes to differentiate tensors while keeping track of the tensorial character of the results. In particular,

$$\nabla \overrightarrow{A} = \overrightarrow{e}^i \frac{\partial}{\partial x^i}(A^j \overrightarrow{e}_j) = \overrightarrow{e}^i \overrightarrow{e}_j \frac{\partial A^j}{\partial x^i} + \overrightarrow{e}^i A^j \Gamma^m_{ji} \overrightarrow{e}_m$$

is a second-rank tensor, with components

$$A^m_{;i} = \frac{\partial A^m}{\partial x^i} + A^j \Gamma^m_{ji}; \tag{7.22}$$

the; notation is used for the covariant differentiation, which is a generalized differentiation that takes the variation of the basis vectors into account. This means that one takes into account the effect of the curved coordinate lines on the differential by the substitution

$$dA^m = \frac{\partial A^m}{\partial x^s} dx^s \rightarrow DA^m = \left(\frac{\partial A^m}{\partial x^s} + \Gamma^m_{is} A^i \right) dx^s. \tag{7.23}$$

For a covariant vector, the rule is

$$dA_m = \frac{\partial A_m}{\partial x^s} dx^s \rightarrow DA_m = \left(\frac{\partial A_m}{\partial x^s} - \Gamma^i_{ms} A_i \right) dx^s. \tag{7.24}$$

The rule extends to the tensors of any rank. If

$$T = T_{ik}\,\vec{e}^{\,i}\vec{e}^{\,k},$$

$$\nabla T = T_{ik;l}\,\vec{e}^{\,i}\vec{e}^{\,k}\vec{e}^{\,l}$$

with

$$T_{ik;l} = \frac{\partial T_{ik}}{\partial x^l} - T_{mk}\Gamma^m_{ll} - T_{im}\Gamma^m_{lk}. \tag{7.25}$$

In addition, in agreement with the general rule for raising and lowering indices, $\Gamma_{i,kl} = g_{im}\Gamma^m_{kl}$ and $\Gamma^m_{kl} = g^{im}\Gamma_{m,kl}$. The derivatives of the metric tensor $g_{jm} = \boldsymbol{e}_j.\boldsymbol{e}_m$ with respect to the coordinates involve the affine connections $\frac{\partial}{\partial x^k}g_{jm} = \vec{e}_{\,j}\frac{\partial}{\partial x^k}\vec{e}_{\,m} + \vec{e}_{\,m}\frac{\partial}{\partial x^k}\vec{e}_{\,j}$, and one finds that

$$\frac{\partial}{\partial x^k}g_{jm} = \Gamma^p_{mk}g_{jp} + \Gamma^p_{jk}g_{pm}.$$

Hence one obtains Ricci's lemma

$$g_{ik;l} = 0. \tag{7.26}$$

However, if the metric tensor is given, such relations can be used to obtain the affine connections. With $(k, j, m) \to (j, m, k)$,

$$\frac{\partial}{\partial x^j}g_{mk} = \Gamma^p_{kj}g_{mp} + \Gamma^p_{mj}g_{pk}.$$

With $(k, j, m) \to (m, k, j)$, instead,

$$\frac{\partial}{\partial x^m}g_{kj} = \Gamma^p_{jm}g_{kp} + \Gamma^p_{km}g_{pj}.$$

The sum of the first two minus the third is

$$\frac{\partial}{\partial x^k}g_{jm} + \frac{\partial}{\partial x^j}g_{mk} - \frac{\partial}{\partial x^m}g_{kj}$$

$$= \Gamma^p_{mk}g_{jp} + \Gamma^p_{jk}g_{pm} + \Gamma^p_{kj}g_{mp} + \Gamma^p_{mj}g_{pk} - (\Gamma^p_{jm}g_{kp} + \Gamma^p_{km}g_{pj})$$

$$= 2g_{pm}\Gamma^p_{jk}.$$

To obtain the connection, multiply by g^{im} and sum over m:

$$\Gamma^i_{jk} = \frac{1}{2}g^{il}\left(\frac{\partial g_{jl}}{\partial x^k} + \frac{\partial g_{lk}}{\partial x^j} - \frac{\partial g_{kj}}{\partial x^l}\right). \tag{7.27}$$

For a curve $x^i(s)$, where s is a parameter like a curvilinear length, the tangent at a point of coordinates x^i has components $\frac{dx^i}{ds}$, and so the tangent at a point of coordinates $x^i + dx^i$ has components $\frac{dx^i}{ds} + \frac{d^2x^i}{ds}$, $i = 1, d$. There is an important special case to consider, i.e., the case of geodesic.

Problem 7.1 Find the non-vanishing Christoffel symbol for spherical geometry in 3 dimensions.

Solution 7.1

$$\Gamma^1_{22} = -r, \Gamma^2_{12} = r^{-1} = \Gamma^3_{13} = \Gamma^3_{31}, \Gamma^1_{33} = -r\sin^2\theta, \Gamma^3_{23} = \cot\theta = \Gamma^3_{32},$$

$$\Gamma^2_{33} = -\sin(\theta)\cos(\theta).$$

7.2 Geodesics

Geodesics or shortest distance paths generalize straight lines to general spaces, and we need an equation for them since they play an important role in next chapter. As a simple starting point, suppose the space is $\mathbb{R}^2$, and the geodesic between two points is a straight line, which is the shortest line between any two points A and B, and its tangent is the same at all of its points and coincides with the straight line itself. Next, we decide to describe the same geodetic by some set of curvilinear coordinates.[1] Now the images of A and B are connected by the image of the straight line, the distance AB remains the same, and the image of the straight line becomes the geodesic between A and B. We can search for it, requiring that the tangent t be the same at $x^i + dx^i$ as in x^i, except that its components change solely because of the change of the local reference according to (7.18). The components of the tangent vector are

$$t_i = \frac{dx_i}{ds}, \tag{7.28}$$

so we require

$$\frac{dt}{ds} = 0 \rightarrow \frac{d(t^i e_i)}{ds} = \frac{dt^i}{ds}e_i + t^i\frac{de_i}{ds} = 0,$$

and inserting $d\overrightarrow{e}_i = \Gamma^k_{ij}dx^j\overrightarrow{e}_k$ which is Eq. (7.16) in the form $\frac{de_i}{ds} = \Gamma^k_{ij}\frac{dx_j}{ds}e_k$, we obtain

$$0 = \frac{dt^i}{ds}e_i + t^i\Gamma^k_{ij}\frac{dx_j}{ds}e_k.$$

[1] The Γ symbols contain the information about the curvature of the space; for this reason, the geodetic equation is the same for curved spaces and for Euclidean spaces described by means of curved coordinates.

Now, reshuffling indices with $k \to i, i \to j, j \to k$ in the second term, we get

$$0 = \left(\frac{dt^i}{ds} + t^j \Gamma^i_{jk} \frac{dx_k}{ds} \right) e_i .$$

Thus, using (7.28), we obtain the equation for the geodesic, namely

$$\frac{d^2 x^i}{ds^2} + \Gamma^i_{jk} \frac{dx^j}{ds} \frac{dx^k}{ds} = 0. \tag{7.29}$$

As noted above, a geodesic has another notable property: it is a length-minimizing curve. This property remains true, and also the above relations continue to hold, even if the base space is curved and has many dimensions.

There is a pleasant and useful way to generate geodesics, which is also a shortcut for evaluating the Γ connections. Consider the function

$$\mathcal{L}_{Geo}(x, \dot{x}) = \frac{1}{2} g_{mn} \dot{x}^m \dot{x}^n, \tag{7.30}$$

where the dots denote derivations with respect to the curvilinear length s. Its dependence on coordinates and *velocities* suggests that it resembles a *Lagrangian*. I am going to show that the Euler–Lagrange equations $\frac{d}{ds} \frac{\partial \mathcal{L}_{Geo}}{\partial \dot{x}^p} - \frac{\partial \mathcal{L}_{Geo}}{\partial x^p} = 0$ coincide with the geodesic equations in covariant form. Indeed,

$$\frac{\partial \mathcal{L}_{Geo}}{\partial \dot{x}^p} = \frac{1}{2} (g_{mn} \delta^m_{np} \dot{x}^n + g_{mn} \delta^n_{np} \dot{x}^m) = g_{pn} \dot{x}^n$$

and

$$\frac{\partial \mathcal{L}_{Geo}}{\partial x^p} = \frac{1}{2} \partial_p g_{mn} \dot{x}^m \dot{x}^n .$$

Thus, one finds the equations of motion

$$\frac{d}{ds} g_{pn} \dot{x}^n = \frac{1}{2} (\partial_p g_{mn}) \dot{x}^m \dot{x}^n . \tag{7.31}$$

Doing the derivative,

$$g_{pn} \ddot{x}^n + \partial_m g_{pn} \dot{x}^m \dot{x}^n - \frac{1}{2} \partial_p g_{mn} \dot{x}^m \dot{x}^n = 0. \tag{7.32}$$

Then, one symmetrizes $\partial_m g_{pn} \dot{x}^m \dot{x}^n$ with respect to the exchange of the dummy indices,

$$\partial_n g_{pm} \dot{x}^m \dot{x}^n \to \frac{1}{2} \partial_n g_{pm} \dot{x}^m \dot{x}^n + \frac{1}{2} \partial_m g_{pn} \dot{x}^m \dot{x}^n .$$

In this way, $\ddot{x}^m + \Gamma^m_{pn} \dot{x}^p \dot{x}^n = 0$, which is the geodesic equation (7.29).

7.3 Curvature Tensor

Up to now, our treatment applies indifferently to curved spaces and to Euclidean spaces in which for some reason we are using curved coordinates. However, we are ready to introduce the Riemann–Christoffel curvature tensor which allows to make the distinction and characterize the curvature; this is a central ingredient of our next chapter on General Relativity. From Eq. (7.33), one understands that in a flat space the parallel transport of any vector along a closed circuit

$$\Delta W_i = \oint \Gamma^k_{il} W_k dx^l \tag{7.33}$$

vanishes.

The difference between the components of a vector at two nearby points is

$$d\overrightarrow{A} = A^k_{;j} dx^j \overrightarrow{e}_k, \tag{7.34}$$

where

$$A^k_{;j} = \left(\frac{dA^k}{dx^j} + A^i \Gamma^k_{ij} \right). \tag{7.35}$$

A vector such that $A^k_{;j} = 0$ does not change if $x_j \to x_j + dx_j$, and so it undergoes parallel transport.

The Riemann–Christoffel curvature tensor R can be obtained as follows. By applying to (7.33) the well-known Stokes theorem

$$\oint_{\delta\Sigma} \overrightarrow{F} d\overrightarrow{l} = \oint_{\Sigma} rot\, \overrightarrow{F} . \overrightarrow{n}\, dS, \tag{7.36}$$

one finds that $2\Delta W_k = R^i_{klm} W_i \Delta f^{lm}$, where Δf^{lm} is the surface element and

$$R^m_{ikl} = \Gamma^n_{il}\Gamma^m_{nk} - \frac{\partial}{\partial x^l}\Gamma^m_{ik} - \Gamma^n_{ik}\Gamma^m_{nl} + \frac{\partial}{\partial x^k}\Gamma^m_{il}. \tag{7.37}$$

In general, for a vector A, one verifies[2] that

$$A_{i;k;l} - A_{i;l;k} = R^m_{ikl} A_m. \tag{7.38}$$

[2] Calculating the L.H.S., one can derive the curvature tensor in an alternative, simpler way; the above derivation using parallel transport over a closed circuit helps more our intuition.

A nonzero R^m_{ikl} implies that the space is curved; it has d^4 components, where d is the dimensionality of the space. There are many symmetry relations among them. The Bianchi identity states that

$$R^m_{ikl;n} + R^m_{ink;l} + R^m_{iln;k} = 0. \tag{7.39}$$

The Ricci tensor is obtained by the following contraction:

$$R_{ab} = R^c_{abc}. \tag{7.40}$$

One finds the symmetric tensor:

$$R_{ik} = \frac{\partial \Gamma^l_{ik}}{\partial x^l} - \frac{\partial \Gamma^l_{il}}{\partial x^k} + \Gamma^l_{ik}\Gamma^m_{lm} - \Gamma^m_{il}\Gamma^l_{km}. \tag{7.41}$$

Moreover, the curvature scalar is defined by

$$R = R^a_a. \tag{7.42}$$

Spaces of constant curvature are of special interest. Let us consider a four-dimensional space with Cartesian coordinates x_i, $i = 1 \ldots 4$, in which we wish to embed a three-dimensional constant curvature subspace. Immediately, one thinks about the three-dimensional surface of a 4-sphere. Let the radius be $a > 0$. The surface is defined by $\sum_i^4 x_i^2 = a^2$, and one can set up spherical coordinates in the three-dimensional surface with $r^2 = \sum_i^3 x_i^2$. Thus, on the surface of the hypersphere,

$$dx_4 = -\frac{x_1 dx_1 + x_2 dx_2 + x_3 dx_3}{\sqrt{a^2 - \sum_i^3 x_i^2}} = \frac{dr^2}{\sqrt{a^2 - r^2}}.$$

The metric in this space is

$$dl^2 = \frac{dr^2}{1 - \frac{r^2}{a^2}} + r^2(d\theta^2 + \sin^2(\theta) d\phi^2). \tag{7.43}$$

This is an isotropic space that has a finite volume and constant curvature. It is customary to put $r = a\sin(\chi)$ with χ in the interval $(0, \pi)$ in order to simplify the first term and bring a in front of the expression; then,

$$dl^2 = a^2[d\chi^2 + \sin^2(\chi)(d\theta^2 + \sin^2(\theta) d\phi^2)]. \tag{7.44}$$

There is, however, another possible three-dimensional constant curvature hypersurface, which is obtained by letting a to be imaginary, that is, replacing a^2 with $-a^2$ throughout. Then, the metric becomes

$$dl^2 = \frac{dr^2}{1 + \frac{r^2}{a^2}} + r^2(d\theta^2 + \sin^2(\theta)d\phi^2). \tag{7.45}$$

In this case, the first term is simplified if we put $r = a\sinh(\chi)$, and we may write

$$dl^2 = a^2[d\chi^2 + \sinh^2(\chi)(d\theta^2 + \sin^2(\theta)d\phi^2)]. \tag{7.46}$$

Chapter 8
Gravity

General Relativity, the most beautiful of theories according to Lev Landau, has been tested and verified in many ways with increasing accuracy. Many of the modern tests were science fiction in 1915, when the theory was published. A series of striking experiments and astronomical observations have demonstrated that it is extremely successful and far reaching, too. Indeed, Einstein expected that many of the predictions could not be tested experimentally; now there is an impressive body of evidence with practical applications in Science, and also in everyday life. This chapter is not a substitute for a full course, but will introduce the reader to the main concepts and to recent developments.

8.1 The Ehrenfest Paradox

The Newton gravity stood (and stands) as one of the major achievements of Science, and before relativity the only known discrepancy with observation was a minor effect on the perihelion of Mercury.[1] In a 1893 paper, Oliver Heaviside developed an analogy between gravity and electromagnetism and suggested that the instantaneous Newton gravity should be an approximation, and some sort of waves should exist.

Hidden in Special Relativity, there are important hints that led Einstein to the right direction for its generalization. A paradox named after Paul Ehrenfest dates back to 1909. He considered an ideal rigid right circular cylinder rotating around

[1] Some people, including epistemologists, have difficulty to understand that the physical theories are not like fashion, are not dismissed, but generalized. The more general theory includes the older one as a special case. There are no paradigms and nothing is thrown away.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_8

its axis. By symmetry, the section of the cylinder must remain circular, and the radius R should not be affected by the motion, since it is always orthogonal to the velocity. But the circumference can be visualized as a polygon with many sides, and they move parallel to the velocity $v = \omega R$. So, Ehrenfest concluded that the length of the circumference in the laboratory frame K should be $2\pi\,\mathrm{R}\sqrt{1-\frac{v^2}{c^2}}$, and this was a striking paradox. The problem was obscured by practical complications: no rigid bodies exist; it is impossible to perform this experiment in the laboratory and attain relativistic speeds. However, Einstein found the key to the problem. The thought experiment must be done correctly. The experimenter in the lab (reference K) divides the radius R and the circumference in a large number N of small rods, say 1 mm long, and measures them by a meter. By a meter one measures proper lengths. Therefore, he cannot really measure the moving platform, but the fixed margin. In this way he finds that the circumference is 2πR, within the experimental error.

Then he goes to the rotating platform (reference K′) bringing the same meter, and repeats the operation. Of course he feels the inertial forces, but the circumference is still a circumference, as is obvious by symmetry. The radius R of the circumference is orthogonal to the motion and is R again. In the reference K′, a length along the circumference is a *proper length*; after dividing the circumference in small rods, say 1 mm long as in K, he measures the circumference again by counting the rods. An observer in the inertial system K also observes them; he finds that they are Lorentz contracted and shorter than 1mm, but they cover the entire moving circumference because they are more numerous than the Ns that cover the stationary circumference. The observer in K′ agrees with the one in K about the number of the small rods, but he measures them to be 1 mm long, so he disagrees about the length of the circumference, concluding that in K′ the Euclidean geometry does not hold. The physical difference between K and K′ is that K is inertial, while in K′, there are inertial forces.

But there is more. The observer in K′ should feel inertial forces that increase with the distance from the center and note that the clocks that are further from the origin run slower than those that are nearer. The conclusion is unavoidable: not only Euclidean rules do not apply to space in the rotating frame, but space–time is affected!

8.2 Principle of Equivalence

We have already seen that Classical Mechanics allows us to choose any reference system and the inertial forces are automatically generated by the Lagrangian formalism in a simple way; therefore, the extension of the theory to include accelerated systems is a logical necessity in the first place. The above example reveals that the inertial forces are related to a more general geometry of space–time, and this is in line with the well-known fact that they produce accelerations (e.g., centrifugal and Coriolis) that are *mass-independent*. An elephant and a mosquito receive the same acceleration from a rotating platform. But this mass-independence of the acceleration has another time-honored, celebrated occurrence, namely, Gravity.

Dropping from the tower of Pisa two spheres of the same radius,[2] but different weight, Galileo allegedly demonstrated experimentally that all bodies have the same acceleration and touch the ground together. This statement is known as the *weak equivalence principle*. Aristotle had stated that the heavier mass must arrive before the light mass. Galileo arrived at a result that corrected a seemingly intuitive millennial error supported by a long indisputable authority, made unquestionable by the Church. Indeed, he guessed the result of experiments that physicists would continue to perform for centuries with increasing accuracy. Eötwös reached an accuracy of $5\ 10^{-9}$ in 1922, and in 2012, Lunar Laser Ranging claimed an accuracy of 10^{-13}. Moreover, the Cavendish 1798 experiment and more recent measurements say that the gravitational constant is

$$G = 6.67 \times 10^{-11} \frac{m^3}{Kg \times s^2}. \tag{8.1}$$

But Einstein realized that he needed a more general equivalence principle than that. He postulated that *the outcome of any non-gravitational* local *experiment in a free-falling laboratory is independent of velocity and location in space*. Local means in a region where fields are uniform. Einstein remarked that this principle implies the complete physical equivalence of a gravitational field and a corresponding acceleration of the reference system, and also that no local experiment can distinguish between gravitational and inertial forces; all such forces have a geometrical nature. This is the *Einstein principle of equivalence*.[3]

The same principle is in operation on a satellite in circular orbit of radius R around some planet or star. If at the center of the circle, there is a star of mass M, the radial force felt by an Astronaut would be expected classically to be

$$F = m(\frac{v^2}{R} - \frac{GM}{R^2}).$$

By suitably selecting v, the force F vanishes, so our Astronaut is in free-fall. This is well known to happen in space laboratories in orbit. Moreover, the Astronaut could note that at smaller distance R, the attraction of the star prevails, while at larger R the centrifugal force wins. Therefore, the vanishing of F is only local, and extended objects are stretched by tidal forces. The fact remains that there is a trade-off between the field of gravity and an accelerated reference system. The equilibrium between forces of different origin is a familiar situation, but the striking fact is that the same situation holds regardless of the mass of the body, since the m

[2] In order to eliminate the effects of the resistance of the air.

[3] The strong equivalence principle states that the outcome of any local experiment (involving Gravity or not) in a free-falling laboratory is independent of velocity *and location*. This requires that the gravitational constant G be the same at any location in space–time. Einstein's General Relativity satisfies the strong equivalence principle, and the gravitational constant G is simply a constant, G. The Brans–Dicke theory, developed in 1961, instead, allows for a position-dependent intensity of gravitation mediated by a scalar field $\phi = \frac{1}{G}$. The astronomers have looked for the possibility that the laws of Nature can be different in far galaxies, but so far the results are consistent with a null result.

that enters the centrifugal term (inertial mass) is the same as the m in the second term (gravitational mass). Therefore, inertial and gravitational forces are essentially the same sort of thing, and we may identify the two masses, although a priori inertial and gravitational masses could be different.

We can already get interesting qualitative results through other thought experiments based on the Principle of equivalence. Consider an observer in a cabin without windows inside a spaceship; he drops a stone from a height of one meter, and finds that it falls downwards with acceleration g, directed, say, along the negative z axis. Suppose g is the same as the acceleration of gravity on Earth. The observer stands normally and does not float. He concludes that the ship probably stands on the ground, or maybe it flies like a plane, without a vertical acceleration. However, that conclusion is not granted. The ship could be in deep space, where the force of gravity is negligible, but subject to an upwards acceleration g. According to the Einstein principle of equivalence, no experiment made inside the cabin can distinguish between the two situations.

Now assume that a light ray with the Poynting vector along the x axis enters horizontally from a small hole in a wall of the cabin, at $z = z_0$. We know from Special Relativity that in any inertial system, independent of the vertical velocity of the cabin, the light propagates in a straight line, producing a spot on the opposite side of the cabin, still at $z = z_0$. If, instead, the ship were accelerated toward the high, the beam should sag and hit the opposite wall at a spot at $z < z_0$. In the time it takes for the light to cross the cabin, the latter has gained speed and advanced more than in the inertial situation. In addition, the light at the spot should be Doppler shifted toward the blue, because the Poynting vector there should have an additional negative component along z. But then, according to Einstein, the light *falls* in a field of gravity; although it has no mass, the ray bends and changes color. These effects are not contained in the Maxwell equations, which are implicity intended for inertial systems. The color change has been verified in the Laboratory, but is of special interest in Astrophysics as the *gravitational redshift* of the light coming from massive compact stars. To make the argument more quantitative, we can go back to the ship accelerated upwards with acceleration g. We set a light source at height h with respect to the detector; during the journey of the light, which lasts a time $\Delta t \sim \frac{h}{c}$, the speed of the spaceship changes by $\Delta v \sim \frac{gh}{c}$; the Doppler blue shift is $\frac{\Delta\nu}{\nu} \sim \frac{\Delta v}{c} \sim \frac{gh}{c^2}$. For any material body of mass m, this ratio could be rewritten as the ratio $\frac{mgh}{mc^2} = \frac{m\Delta\Phi}{mc^2}$ of the change in gravitational potential energy $m\ gh = m\Delta\Phi$ to the rest energy. $\Delta\Phi$ is the drop in gravitational potential. Here, the Principle of Equivalence enters. In a field of gravity,

$$\frac{\Delta\nu}{\nu} \sim \frac{\Delta\Phi}{c^2}. \tag{8.2}$$

The light *falling* toward the star shifts toward the blue, but the light reaching us from the star reddens.[4] But in order to achieve a coherent formulation of the Principle of Equivalence, we need the Principle of General Covariance.

8.3 The Principle of General Covariance and the Curved Space–Time

In accelerated frames and/or in the presence of gravity, the space–time is no longer flat, so we must be able to work with curvilinear coordinates x^μ, $\mu = 0, 1, 2, 3$ and to cope with general changes of frame to new coordinates x'^μ; the transformation needs to be invertible and differentiable, for physical reasons, but otherwise is general. The *principle of General Covariance* (also known as the *principle of General Relativity*) states that the laws of physics must be the same equations for *all* observers, inertial or not. Removing the inertial frames from their exclusive role in Special Relativity makes things much more complicated but was logically necessary. We have noted that already the Lagrangian formulation of Classical Mechanics allows to work in any reference and leads to a much deeper understanding of the theory.

In Special Relativity, the infinitesimal interval between two events is, according to Eq. (6.27),

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu, \tag{8.3}$$

with $\eta_{\mu\nu} = 0$ for $\mu \neq \nu$ and $\eta_{00} = -1, \eta_{11} = \eta_{22} = \eta_{33} = 1$; this means that the geometry is pseudo-Euclidean and the space–time is flat.[5] A change of inertial reference may be done by combining the special Lorentz transformation

$$x'_0 = \gamma(x_0 - \beta x_1), x'_1 = \gamma(x_1 - \beta x_0), x'_2 = x_2, x'_3 = x_3$$

with translations and rotations. We must generalize Eq. (8.3) by introducing a symmetric[6] *metric tensor* with covariant components $g_{\mu\nu}$ by setting

$$\boxed{-ds^2 = g_{\mu\nu} dx^\mu dx^\nu} \tag{8.4}$$

For an inertial system $g_{\mu\nu} = diag(-1, 1, 1, 1)$. From $dx'^\lambda = \frac{\partial x'^\lambda}{\partial x^\mu} dx^\mu$, $x'^\rho = \frac{\partial x'^\rho}{\partial x^\nu} dx^\nu$, one can deduce how the covariant tensor components must transform

[4] Again, this implies that clock go slower in a field of gravity, as they do under the action of a centrifugal force.

[5] I adopt the convention used by Landau–Lifshitz, Stephani and probably the majority of Authors. Moreover, the Greek indices run from 0 to 3 and the Latin indices from 1 to 3.

[6] g is a symmetric tensor, $g_{\mu\nu} = g_{\nu\mu}$, while the electromagnetic tensor F is antisymmetric. In the unified field theory proposed by Einstein and Schrödinger in the 50 s the field is described by a tensor with a symmetric gravitational part and antisymmetric electromagnetic contribution.

$$g_{\mu\nu} = g'_{\lambda\rho} \frac{\partial x'^{\lambda}}{\partial x^{\mu}} \frac{\partial x'^{\rho}}{\partial x^{\nu}}. \tag{8.5}$$

Consider a reference where two events take place at the same point and almost at the same time. The observer describes the facts in terms of Eq. (8.4), with $dx^i = 0$ for i $= 1, 2, 3$; the infinitesimal interval ds is given by $ds^2 = -c^2 d\tau^2 = g_{00} dx_0^2$, where τ is the proper time, and so

$$c d\tau = \sqrt{-g_{00}} dx^0. \tag{8.6}$$

Clocks show their own proper time τ, not the coordinate time x_0. However from a calculated dx_0, one can obtain the physical $d\tau$ if g_{00} is known. Equation (8.5) is the covariant transformation law for the components tensors with two indices. The contravariant components (see Chap. 7) transform according to the rule

$$g^{\mu\nu} = g'^{\lambda\rho} \frac{\partial x^{\mu}}{\partial x'^{\lambda}} \frac{\partial x^{\nu}}{\partial x'^{\rho}}. \tag{8.7}$$

As in Special Relativity, a central role is played by a variety of functions of the space–time point P that remain invariant under a change of frame. Scalars, 4-vectors and more general objects with many components that transform in this way are called covariant tensors. Their components vary in such a way as to compensate for the frame change; some tensors generalize those of Special Relativity, but new tensors are also needed. We shall use the essential formalism summarized in Chap. 7.

8.3.1 Space Geometry in Stationary Problems

Can we define a space metric at a given time and make an instant map, say, of a region of the Universe? No. In general, space and time are mixed in such a way that one cannot do that. This is not a mere mathematical problem, but the consequence of the fact that measuring distances takes time, and so distances change before one can measure them, and vary with the path of the signals used to measure them. In Special Relativity, in order to measure the distance to a far object, one can use the exchange of light signals and the fact that the proper time the light takes to cover any distance is zero. However, in the presence of gravity, this is not possible in principle. This is a fundamental complication when we try to define the distance of a galaxy. Apart from the obvious practical impossibility to exchange signals with remote objects like those considered in Astrophysics, a further trouble is that the light can follow different routes while the Universe expands and the galaxies move.

However, some interesting problems are stationary, and so we can (so to speak) separate time from space. Consider two nearby points A and B separated by an infinitesimal vector $d\vec{r} = (dx_1, dx_2, dx_3)$. In order to determine their distance, imagine we can send a signal from B to A and receive the immediate response signal from A to B. The time between the two signals, after multiplication by $c/2$, gives the

distance AB. Now, we impose $ds^2 = 0$ (as appropriate to the propagation of light), use $-ds^2 = g_{\mu\nu}dx^\mu dx^\nu$ and find two roots corresponding to the coordinate times:

$$dx^{0-} = \frac{g_{0k}dx^k - \sqrt{(g_{0i}g_{0k} - g_{00}g_{ik})dx^i dx^k}}{-g_{00}}, i, k = 1, 2, 3, \tag{8.8}$$

$$dx^{0+} = \frac{g_{0k}dx^k + \sqrt{(g_{0i}g_{0k} - g_{00}g_{ik})dx^i dx^k}}{-g_{00}}, i, k = 1, 2, 3. \tag{8.9}$$

Why two times? Because Eq. (8.10) refers to both signals, from B to A and back, and it is the difference between these coordinate times that gives us the coordinate time taken by the exchange of signals.

$$dx^{0-} - dx^{0+} = \frac{-2\sqrt{(g_{0i}g_{0k} - g_{00}g_{ik})dx^i dx^k}}{g_{00}}, i, k = 1, 2, 3. \tag{8.10}$$

We have found the duration of the experiment in coordinate time x^0; to obtain the proper time we must recall that Eq. (8.6) tells us to include a factor $\sqrt{-g_{00}}$. Therefore,

$$d\tau = -2\sqrt{(g_{\alpha\beta} - \frac{g_{0\alpha}g_{0\beta}}{g_{00}})dx_\alpha dx_\beta}. \tag{8.11}$$

Now, multiplication by $c/2$ gives the distance AB.

Thus, in the case of stationary problems, when $g_{\mu\nu}$ is time-independent, one can define the 3×3 metric tensor of space γ_{ij}, such that the element of distance is

$$dl^2 = \gamma_{ij}dx^i dx^j, \tag{8.12}$$

where i, j = 1, 2, 3. It turns out that

$$\gamma_{ij} = g_{ij} - \frac{g_{0i}g_{0j}}{g_{00}}. \tag{8.13}$$

8.3.1.1 Curved Space–Time in a Rotating Frame and the Ehrenfest Paradox

A simple example is useful to build an insight into the components of the metric tensor and how the information is distributed therein. Recall the transformation (2.32) from an inertial system with Cartesian coordinates (x_0, y_0) to a rotating platform with Cartesian coordinates (x, y) in Classical Mechanics. The inverse transformation, letting the two systems coincide for $t = 0$, is

$$x_0 = x\cos(\omega t) + y\sin(\omega t),\ y_0 = x\cos(\omega t) + y\sin(\omega t). \tag{8.14}$$

In the inertial system $ds^2 = c^2dt_0^2 - dx_0^2 - dy_0^2 - dz_0^2$. We transform to the rotating system using Eq. (8.14) and letting ω=2$\pi\nu$. After some algebra one finds

$$ds^2 = [c^2 - \omega^2(x^2 + y^2)]dt^2 - 2\omega(ydx - xdy)dt - dx^2 - dy^2 - dz^2. \quad (8.15)$$

To avoid unnecessary complications at this stage, since Eq. (8.15) neglects all the relativistic corrections, we assume that the platform is small enough and/or rotates slowly enough that the coefficient of dt^2 does not change sign. Therefore, we attempt to introduce the corrections with respect to special relativity by setting

$$g_{\mu\nu} = -\eta_{\mu\nu} + h_{\mu\nu}, \quad (8.16)$$

where $\eta = diag(-1, 1, 1, 1)$ and $h_{\mu\nu}$ are the elements of the matrix

$$h = \begin{pmatrix} -\frac{\omega^2}{c^2}(x^2 + y^2) & -\frac{\omega y}{c} & \frac{\omega x}{c} & 0 \\ -\frac{\omega y}{c} & 0 & 0 & 0 \\ \frac{\omega x}{c} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (8.17)$$

This simple example shows some features that we shall encounter later: the classical potential (centrifugal in this case) is in h_{00}, hence also in g_{00}, while h_{0i} shows the mixing of time with x and y that is related to the velocity-dependent Coriolis force. In this case, from the above result (8.12, 8.13), one finds the distance element

$$dl^2 = d\rho^2 + dz^2 + \frac{\rho^2 d\phi^2}{1 - \omega^2\frac{\rho^2}{c^2}}, \quad (8.18)$$

$\rho = (x, y)$. Setting $dz = d\rho = 0$, one finds[7] $\oint dl = \frac{2\pi\rho}{\sqrt{1-(\frac{\omega\rho}{c})^2}}$.

So, an observer at rest in the rotating system would find that a circumference centered on the origin of the plane z = 0 is longer than $2\pi r$ by a factor $\frac{1}{\sqrt{1-\omega^2\frac{r^2}{c^2}}}$, as stated earlier by a thought experiment about the Ehrenfest paradox.[8]

8.3.1.2 Synchronization, GPS and Coordinated Universal Time (UTC)

Having set a reference system, it is crucial to be able to synchronize all the clocks; only in this way one can tell which of two events A and B is earlier in that reference.

[7] A paper by F. Goy and F. Selleri, arXiv:gr-qc/9702055 remarks that $R_{ijkl} = 0$, hence the space–time is flat; however, we shall see below that in the rotating reference, the space is not flat.

[8] For an alternative formalism see the interesting paper by Matteo Luca Ruggiero (2003).

However, we have seen that clocks in General Relativity run slower near important masses. If we wish to synchronize the clocks in two different points, we must exchange a signal and consider that the instant in which the signal is received as simultaneous with the mean of the times of emission and return of the signal. Then, using Eq. (8.10), one finds

$$dx^0 = -\frac{\sum_\alpha g_{0\alpha} dx^\alpha}{g_{00}}. \tag{8.19}$$

This does not imply that the synchronization is impossible in general. For example, in the Rindler problem (Chap. 6), the crew experiences an effective gravitational field and the clocks in the ground run slowest, but it is still possible to compensate for that. In other terms, gravity causes lower clocks to run slower. No problem, we can adjust these clocks to go faster. So all the members of the crew can agree which of two event is earlier. Right?

Indeed, this works since the interval (6.12) has no off-diagonal term, but does not work if there is rotation. Then, in a closed circuit, one cumulates a time difference

$$\Delta x^0 = -\oint \frac{\sum_\alpha g_{0\alpha} dx^\alpha}{g_{00}}. \tag{8.20}$$

Therefore, it is evident that it is impossible to synchronize clocks in a rotating body. This is quite serious, because if observers cannot agree about the order of events in time, the very notion of a reference loses meaning. Since the Earth rotates, and artificial satellites also rotate, this concerns us. Putting numbers is, it turns out that the discrepancies are of the order of tens of nanoseconds. However, even very short times are significant for the Global Positioning System (GPS) which defines the Greenwich Mean Time and must also take into account several other effects of special and general relativity. The "Earth-Centered Earth-Fixed frame" (ECEF) in which we live is not appropriate. In order to define a proper time with the necessary precision, one defines the "Earth-Centered Inertial Frame" (ECI) in which Earth spins; this admits a valid time which is called the Coordinated Universal Time (UTC); however, this is not the actual time of our life.

8.3.1.3 Rotating Platform and Sagnac Effect

Here we revisit the Sagnac effect (Chap. 6) from the standpoint of General Relativity; to this end, we write the interval in the rotating platform using cylindrical coordinates:

$$ds^2 = (c^2 - r^2\omega^2)dt^2 - 2\omega r^2 d\phi dt - dr^2 - r^2 d\phi^2 - dz^2. \tag{8.21}$$

Since the light rays ($ds = 0$) are circular, we set $dr = 0, dz = 0$ and find

$$dt = \pm \frac{R d\phi}{c^2 - R^2\omega^2} \tag{8.22}$$

which leads directly to the delay holding in the laboratory frame; however, this is difficult to understand, since this should be the time measured in the rotating frame. For this reason, the relativistic interpretation of the Sagnac effect is still controversial.[9]

8.3.2 Generalized Equation of Motion

The four-velocity of a particle in special Relativity has components

$$v_i = \frac{dx^i}{d\tau}, \tag{8.23}$$

where the proper time τ is linked to te interval s by $s = c\tau$. It is clear physically that the velocity cannot be anything else, therefore in General Relativity, we keep this expression. The four-acceleration changes, however. To satisfy the general covariance requirement, all quantities appearing in the fundamental equations must be tensors and all the derivatives must be replaced with covariant derivatives of the type (7.19), that is, $d \to D$. Therefore, the equation of motion of a test mass m subject to a force field of electromagnetic origin f must become

$$m\left(\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\nu\sigma}\frac{dx^\nu}{d\tau}\frac{dx^\sigma}{d\tau}\right) = f^\mu, \tag{8.24}$$

Therefore, a free-falling mass (in the absence of f) follows a geodetic in space–time. The familiar parabola is actually a geodetic, and although the curvature of our ambient space is very small, the curvature of space–time is clearly visible. Now it is obvious that the term involving the connection coefficients stands for effective inertial and gravitational forces.

This allows us to figure out a physical and dynamical significance to the metric tensor and its derivatives, at least in problems in which the relativistic effects are small. As in the last subsection, one starts by writing

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{8.25}$$

where $\eta_{\mu\nu}$ is the limit of Special Relativity and $h_{\mu\nu}$ is a correction that is considered small neglecting its powers. Now we take advantage of the observation made earlier that the non-relativistic potential resides in the g^{00} element; this observation leads quickly to the correct result. Let us reproduce Eq. (7.27)

$$\Gamma^k_{ij} = \frac{1}{2}g^{im}\left(\frac{\partial g_{jm}}{\partial x^k} + \frac{\partial g_{mk}}{\partial x^j} - \frac{\partial g_{kj}}{\partial x^m}\right).$$

[9] Bhadra et al. (2022).

The summation over m produces 12 terms. Neglecting the Coriolis-like off-diagonal g elements, one finds

$$2\Gamma^k_{ij} \sim g^{00}\frac{\partial g^{00}}{\partial x^k}\delta_{i0}\delta_{j0} + g^{00}\frac{\partial g^{00}}{\partial x^j}\delta_{i0}\delta_{k0} - g^{ii}\frac{\partial g^{00}}{\partial x^i}\delta_{k0}\delta_{j0},$$

which yields

$$\Gamma^k_{00} \sim \frac{1}{2}g^{00}\frac{\partial g_{00}}{\partial x^k}, \Gamma^0_{0j} \sim \frac{1}{2}g^{00}\frac{\partial g_{00}}{\partial x^j}, \Gamma^0_{i0} \sim -\frac{1}{2}g^{ii}\frac{\partial g_{00}}{\partial x^i}.$$

Now we are interested in the equation of motion ($\mu = k = 1, 2, 3$), so we use Γ^k_{00}, which we may further approximate as $\Gamma^k_{00} \sim \frac{1}{2}\frac{\partial h^{00}}{\partial x^k}$. It turns out that the equations of motion are approximated by

$$m\frac{d^2x^i}{dt^2} = -m\frac{\partial}{\partial x_i}\frac{mc^2}{2}h_{00}. \tag{8.26}$$

This is the Newtonian result and confirms that the 00 element of the metric tensor in the non-relativistic limit is

$$g_{00} = 1 + \frac{2V}{c^2}, \tag{8.27}$$

where V is the gravitational potential in the weak field case. Through the gravitational redshift, one can measure the gravity on the surface of stars. Currently, the precision in measurement of time is so good that one can appreciate 10^{-18} seconds, and the difference in gravity between different floors of a building can be measured.

8.3.3 Generalized Maxwell Equations

The electromagnetic tensor

$$F_{\mu\nu} = A_{\nu;\mu} - A_{\mu;\nu} = \frac{\partial A_\nu}{\partial x_\mu} - \frac{\partial A_\mu}{\partial x_\nu} \tag{8.28}$$

keeps its form because the terms involving the Γ symbols cancel each other; the first couple of Maxwell's equations also keeps the same form as that in Special Relativity. However, the second pair of equations must be written as

$$F^{\mu\nu}_{;\nu} = \frac{4\pi}{c}J^\mu. \tag{8.29}$$

A light ray with wave vector k^μ goes along a straight line $dk^\mu = 0$ if there are no gravity fields; in a field, this generalizes to $Dk^\mu = 0$. In other words, the ray follows a null geodetic,

$$\frac{dk^\mu}{dx^\beta} + \Gamma^\mu_{\alpha\beta} k^\alpha = 0. \tag{8.30}$$

8.4 Einstein Field Equations

It was clear to Einstein from the start that Newton's theory had to be his non-relativistic limit. Newton's gravitational potential V satisfies the Poisson equation

$$\nabla^2 V = 4\pi G \rho(\boldsymbol{r}, t), \tag{8.31}$$

where $\rho(\boldsymbol{r}, t)$ is the matter density, which is a scalar in the non-relativistic theory, and $G = 6.67408^{-11} m^3 kg^{-1} s^{-2}$. Newton's scheme allowed for an immense conceptual progress and an accurate description of the working of the solar system.[10] However, it is not tenable in the light of the Principles of General Relativity and Equivalence.

Before Relativity, Newton's theory was in complete agreement with the data, with the sole exception of a tiny secular shift in Mercury's periphery, which remained after introducing all the corrections due to the motions of the other planets. But someone noticed the similarity with electrostatics, and Oliver Heaviside suggested that the Poisson equation should be supplemented by some sort of gravito-magnetic field in order to allow for a finite speed of propagation of the gravitational interactions.

We must summarize the main results of a research that engaged Einstein for quite a long time, namely, the field equations that replace the Newton Eq. (8.31). A general theory cannot be deduced from a special theory, so he had to choose the best conjecture. Besides the two Principles, Einstein considered criteria of simplicity and elegance; he had a unique sense of Physics.

The mass density ρ is not available as a source term, because it is not a relativistic invariant, and must be replaced with the energy–momentum tensor $T^{\mu\nu}$ of Eq. (6.38); then, at the left-hand side, we must put a tensor with two indices. Moreover, we have seen in Chap. 6 that in Special Relativity the conservation of energy and momentum requires that $T^{\mu\nu}$ be divergenceless; it is natural to assume that $T^{\mu\nu}{}_{;\nu} = 0$. The metric tensor does not have the right meaning, and the Riemann–Cristoffel tensor (7.37) has too many indices. It can be verified through lengthy algebra that the Einstein tensor

$$G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R \tag{8.32}$$

[10] This is a textbook example of how science proceeds by successive generalizations and not by fashions, revolutions or paradigms as some epistemologists say. Einstein adds new understanding to what Newton had brilliantly understood.

has the property $G^{\mu\nu}_{;\nu} = 0$, so it can replace the L.H.S. of Eq. (8.31). The simple-looking Einstein equations are

$$\boxed{G^{\mu\nu} = kT^{\mu\nu}.} \tag{8.33}$$

These are coupled, nonlinear differential equations for the components of the metric tensor.[11]

The independent Eqs. (8.33) are 10, because $G^{\mu\nu}$ is a symmetric tensor. The left-hand side contains space–time geometry, while the right-hand side describes the sources of gravitation, i.e., matter and fields. Having found the Einstein equations, we must still determine the constant k. This can be done by considering the weak field case, when the relativistic theory must give back the Newton theory. Again, we highlight the corrections with respect to special relativity by setting $g_{\mu,\nu} = \eta_{\mu,\nu} + h_{\mu,\nu}$ like in Eq. (8.16); from the example of the rotating frame, we know that $c^2 h_{0,0}$ corresponds to the non-relativistic potential. More details about the weak field case are deferred to the next section. Here I point out that after long calculations, one finds

$$k = \frac{8\pi G}{c^4}. \tag{8.35}$$

In 1917, Einstein introduced a new term involving the so-called cosmological constant Λ as follows:

$$\boxed{G^{\mu\nu} - \frac{1}{2} g^{\mu\nu}\Lambda = kT^{\mu\nu}.} \tag{8.36}$$

A positive constant Λ produces a sort of antigravity opposing the collapse of the Universe, that should otherwise collapse, and keeping it in equilibrium, as Einstein thought it should be. It can be interpreted as the energy density of the vacuum of space. But in 1929 Hubble discovered that the Universe was actually expanding and Einstein repudiated the Λ term. For many years most people thought that the expansion should be slowed down by gravity, but at the turn of the century the dimming of far supernovae was taken as evidence that the expansion is actually accelerating. Further evidence arose from galaxy surveys and from the cosmic microwave background. Then, Λ was restored, even if nobody knows how this antigravity should arise. In order to agree with the data, Λ must be so small as to be completely immaterial when dealing with the Solar System and its surroundings, yet when the large-scale Universe is considered, it should be a manifestation of the Dark Energy that is thought to overweight all matter (baryonic, dark or whatever else). Attempts to calculate it by Quantum Field Theory turn out to be wrong by many orders of magnitude, and

[11] The equations in the absence of matter can be obtained from the Einstein–Hilbert action

$$S = \frac{1}{2k} \int \sqrt{-g} R d^4 x, \tag{8.34}$$

where $k = \frac{8\pi G}{c^4}$ and $g = det(g_{\mu\nu})$.

this is probably the most spectacular failure of Theoretical Physics. However, we shall put $\Lambda = 0$ since this is appropriate for most purposes.

Lowering an index in (8.33) with $\Lambda = 0$, one finds

$$R^{\mu}_{\nu} - \frac{1}{2} g^{\mu}_{\nu} R = 8\pi G_N T^{\mu}_{\nu};$$

since $g^{\mu}_{\nu} = \delta^{\mu}_{\nu}$, one finds $R = -8\pi G_N T$ where $T = T^{\mu}_{\mu}$. Hence the Einstein equations with $\Lambda = 0$ can also be rewritten in the alternative form

$$R^{\mu\nu} = 8\pi G_N \left(T^{\mu\nu} - \frac{1}{2} T g^{\mu\nu} \right) \tag{8.37}$$

where $T = T^{\mu}_{\mu}$. I stress that these equations are very general since they apply for any distribution of masses however large and fast moving, observed in a reference system that can rotate and be subject to forces; meanwhile the observer is free to adopt any arbitrary curvilinear space–time set of coordinates.

8.4.1 Weak Fields and Linearized Field Equations

The field equations are highly nonlinear. It must be so, since the field is energy and is itself a source of the field; by contrast, the Maxwell equations are linear, since the field is neutral. But nonlinear equations are hard to solve by hand. Only for a few, symmetric situations have analytic solutions been found. In most cases, one must resort to numerical methods. If the field is weak, $h_{\mu\nu}$ is small in Eq. (8.25), and it is possible to linearize and gain some physical insight in general. Linearizing, one finds that

$$\Gamma^{\beta}_{\lambda\alpha} = \frac{1}{2} \eta^{\beta\mu} \left(\frac{\partial h_{\mu\alpha}}{\partial x^{\lambda}} + \frac{\partial h_{\mu\lambda}}{\partial x^{\alpha}} - \frac{\partial h_{\lambda\alpha}}{\partial x^{\mu}} \right). \tag{8.38}$$

Since the h corrections to the metric become potentials in the weak field case, the Christoffel symbol is a combination of field components. Then, using η instead of g to raise and lower indices, one finds through tedious algebra that the Ricci tensor is approximated by

$$R_{\beta\mu} = \frac{1}{2} \left[\eta^{\alpha\lambda} \frac{\partial^2 h_{\beta\mu}}{\partial x^{\alpha} \partial x^{\lambda}} + \frac{\partial^2 h^{\alpha}_{\alpha}}{\partial x^{\mu} \partial x^{\beta}} - \frac{\partial^2 h^{\lambda}_{\beta}}{\partial x^{\mu} \partial x^{\lambda}} - \frac{\partial^2 h^{\alpha}_{\mu}}{\partial x^{\alpha} \partial x^{\beta}} \right]. \tag{8.39}$$

At this point, the reader could expect to find a wave equation yielding the gravitational waves. Since their detection in 2015, the newspapers published the news and wrote about the unusual tensorial character of these waves, so unlike the electromagnetic ones. But life is not so easy. The equation for the propagation in vacuo, namely,

$$R_{ik} = 0, \tag{8.40}$$

is not of the form of a wave equation, but is awfully complicated. This complexity essentially depends on our complete freedom in General Relativity to choose the coordinate system. The system might move about and rotate in an arbitrary way; in addition, there is a gauge invariance analogous to the electromagnetic one. In fact, one can show that by choosing the gauge due to Vladimir Aleksandrovich Fock:

$$\frac{\partial h^{\nu}_{\sigma}}{\partial x^{\nu}} - \frac{1}{2}\frac{\partial h}{\partial x^{\sigma} s} = 0, \tag{8.41}$$

where $h = h^{\alpha}_{\alpha}$, we obtain simply

$$R_{\mu\nu} = -\frac{1}{2}\Box h_{\mu\nu} \tag{8.42}$$

for the Ricci tensor in the limit of weak fields. Then, the Einstein equations reduce to

$$\Box h_{\mu\nu} = -\frac{16\pi G}{c^4}\left(T_{\mu\nu} - \frac{1}{2}T g_{\mu\nu}\right). \tag{8.43}$$

Gravity waves will be discussed in Sect. 8.9.

In the case of low velocities and weak fields, only the only nonzero component of the energy–momentum tensor is $T^0_0 \sim \rho c^2$, and one finds Poisson's equation for the 00 component and for the diagonal spatial components; consequently, in terms of the potential V of the Newton theory, the first correction for the interval is given by

$$ds^2 = -\left(1 + \frac{2V}{c^2}\right)c^2 dt^2 + \left(1 - \frac{2V}{c^2}\right)d\boldsymbol{r}^2. \tag{8.44}$$

8.5 Gravitational Collapse: True Facts, Conjectures and Myths

8.5.1 Schwarzschild Solution and Black Holes

Already in the late eighteenth century, Pierre Simon and Marquis de Laplace conjectured that stars having large mass and small radius could have an escape velocity larger than c, and therefore should be black and invisible. Let us see how black holes emerge from Relativity. The length

$$r_S = \frac{2MG}{c^2} \tag{8.45}$$

is called the Schwarzschild radius for a body having mass M. The ratio $\frac{m\Delta\Phi}{mc^2}$ of the gravitational energy to the rest energy is always small in laboratory experiments but becomes of order unity in the case of black holes, the large collapsing stars that we shall meet soon. Any mass M becomes a black hole if confined to within a Schwarzschild radius $r_S = \frac{2GM}{c^2}$. This is very small compared to the actual size of most objects. See the following table, where the masses are in Kilograms and lengths in meters.

Object	Mass	Radius R	r_S	$\frac{r_S}{R}$
Nucleus	10^{-26}	10^{-15}	10^{-53}	10^{-38}
Atom	10^{-26}	10^{-10}	10^{-53}	10^{-43}
Earth	$6\ 10^{24}$	$6\ 10^{6}$	$9\ 10^{-3}$	10^{-9}
White dwarf	$2\ 10^{30}$	10^{7}	$3\ 10^{3}$	$3\ 10^{-4}$
neutron star	$2\ 10^{30}$	10^{4}	$3\ 10^{3}$	$3\ 10^{-1}$
Sun	$2\ 10^{30}$	$7\ 10^{8}$	$3\ 10^{3}$	10^{-6}
Galaxy	10^{41}	10^{21}	10^{14}	10^{-7}

In 1916, at the German–Russian front, Karl Schwarzschild found the static solution[12] to the Einstein equations for the gravity field of a point-like spherical body *in vacuum*.

It should be stressed that *Schwarzschild did not discover any gravitational singularity at the center of the black hole*. Actually a singularity is a problematic statement in Physics. Rather, he assumed a point-like source of gravity, presumably inspired by the classical result that the field of a spherically symmetric distribution of matter is the same as the field of a point source. This was correct: later it was shown (Birkhoff's theorem) that this is also true in General Relativity.

The most general spherically symmetric metric by symmetry must have the form (in obvious notation)

$$ds^2 = g_{00}(t,r)c^2dt^2 + g_{11}(t,r)c^2dr^2 - r^2(d\theta^2 + \sin^2(\theta)d\phi^2). \tag{8.46}$$

We are using the time-like convention for the signs of the metric coefficients, so $ds^2 > 0$ is a time-like interval and $ds^2 < 0$ is spacelike.

The Schwarzschild metric reads as[13]

$$ds^2 = \left(1 - \frac{r_S}{r}\right)c^2dt^2 - \frac{dr^2}{1 - \frac{r_S}{r}} - r^2d\theta^2 - r^2\sin^2(\theta)d\phi^2. \tag{8.47}$$

[12] The solution was soon extended to the case of a charged non-rotating body (Reissner–Nordström metric).

[13] The derivation of the metric from the field equations is rather long and is available in advanced books like "The theory of Gravitational Interactions", by Maurizio Gasperini, Unitext for Physics, Springer, or "Relativity" by Wolfgang Rindler, Oxford University Press.

Here, $ds = cd\tau$, where τ is the proper time recorded by a clock placed at the chronotopic point specified as (τ, r, θ, ϕ). Then, what is the *coordinate time* t? The relation to the proper time is obtained by fixing the point, that is, by setting $dr = 0, d\theta = d\phi = 0$ in (8.47). One obtains

$$dt = \frac{d\tau}{\sqrt{1 - \frac{r_S}{r}}}. \tag{8.48}$$

For small $\frac{r_S}{r}$ the interval is the same as in special Relativity; therefore, the angles have the usual meaning and the time t is measured by a far observer. At a fixed point $(dr = 0, d\theta = d\phi = 0)$ $dt > d\tau$. This means that the time runs slower near a mass, and spectral lines emitted by atoms in stars appear redshifted to a far observer. We shall say more on this below.

However, at $r = r_S$, we see that $\frac{dt}{d\tau} \to \infty$. This means that there is a radius of infinite redshift or infinite slow motion, corresponding to a so-called *horizon of events*, and a far observer cannot see what happens there or inside. This happens because no signal can overcome the gravity. Conversely then, an infalling observer approaching the horizon must see everything in accelerated motion and touch the horizon, so to say, at the end of the Universe.

Therefore, the Schwarzschild metric is valid only for $r > r_S$.

In order to understand the physical implications of the 4-d metric (8.47), we must use it to model thought experiments in which some observer measures times and lengths in terms of their own proper times and lengths.

Now it is clear that for $r \gg r_S$, the coordinate time t is close to the proper time τ, but in General Relativity, one is interested in the departures from the Newton theory and in situations when such departures may be substantial.

In the same way, r cannot be identified with the distance from the center, since the metric blows up at $r = r_S$ and so the center is off limits.

Moreover, setting $d\theta = d\phi = 0$ and $dt = 0$ in (8.47), we get the equal-times interval, which is minus the square distance $dR^2 = ds^2 = \frac{dr^2}{1 - \frac{r_S}{r}}$; so, the infinitesimal proper length is $dR = \frac{dr}{\sqrt{1 - \frac{r_S}{r}}}$. Again, this diverges as $r \to r_S$. The distance between two r values, say, r_1 and r_2, is given by the integral

$$d[r_1, r_2, r_S] = \int_{r_1}^{r_2} \frac{rdr}{\sqrt{1 - \frac{r_S}{r}}}. \tag{8.49}$$

This integral is elementary and the antiderivative:

$$d = \int \frac{rdr}{\sqrt{1 - \frac{r_S}{r}}} = \sqrt{r(r - r_S)} + r_S \ln[\sqrt{r} + \sqrt{r - r_S}], \tag{8.50}$$

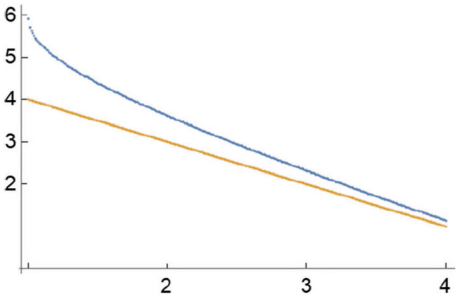


Fig. 8.1 The integral $d[r_1, r_2, r_S = 1]$ represents the coordinate distance between an observer at $r_2 = 5$ (in unit of r_S) and a reckless companion at r_1. Neglecting relativistic effect, the distance should increase linearly and approach $4r_S$ when the companion approaches the horizon. Instead the relativistic distance is given by the upper curve

represents the coordinate distance, in units of r_S, between an observer at $r_2 = 5$ and a reckless companion at r_1. In a flat geometry, when the companion approaches the horizon, the distance should approach $4r_S$. Instead, the relativistic trend is represented in the upper curve and is significantly different (see Fig. 8.1).

8.5.2 *Kruskal and Szekeres Coordinates*

The validity of the metric (8.47) ends at the horizon, but no physical singularity occurs there. An observer right at the horizon of a large black hole would quickly see the end of the universe but apart from that would consider it a place like any other. A coordinate change discovered by Martin Kruskal and György Szekeres allows us to extend the validity to all space–time (except for the central point mass).

The transformation $(r, t) \to (u, v)$ is illustrated in Fig. 8.2. The $u - v$ plane is divided into four regions by the lines $u = \pm v$. Region 1 represents the region outside the horizon and contains the positive u axis, while Region 2 represents the region inside the horizon and contains the positive v axis. The remaining half plane is not used by the above parametrization. The new coordinates u, v are defined in Region 1, that is, for $r > r_S$, by

$$u = \sqrt{\frac{r}{r_S - 1}} \exp\left(\frac{r}{2r_s}\right) \cosh\left(\frac{ct}{2r_S}\right) \tag{8.51}$$

$$v = \sqrt{\frac{r}{r_S - 1}} \exp\left(\frac{r}{2r_s}\right) \sinh\left(\frac{ct}{2r_S}\right), \tag{8.52}$$

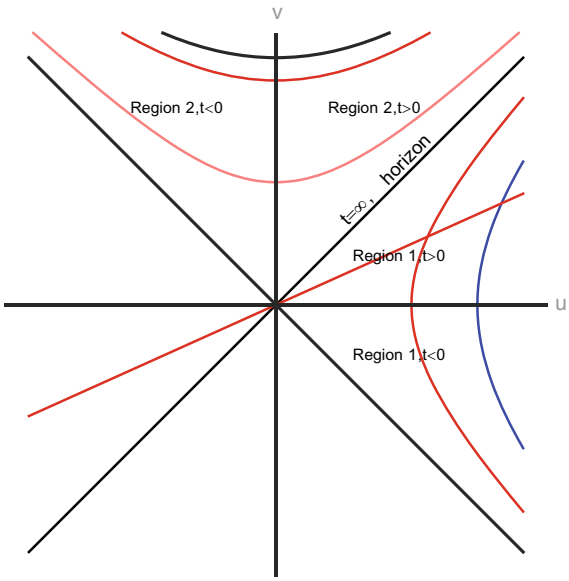


Fig. 8.2 The Kruskal diagram. Region 1 is split by the u axis into $t > 0$ above and $t < 0$ below. Region 1 is split by the v axis into $t > 0$ (right) and $t < 0$ (left). The straight border between the two regions corresponds to $r = r_S$ and $t = \infty$. In Region 1, where we also belong, the hyperbolas $v = \pm\sqrt{u^2 + 1 - \frac{2}{r_S}}$ are constant-r curves, the red one has $r = 1.1r_S$ and the blue one has $r = 1.2r_S$. The straight lines have constant t. In Region 2, which is the physical region, the black hyperbola $v = \sqrt{1 + u^2}$ represents the singularity $r = 0$, the red one has $r = 0.5r_S$ and the pink one $r = 0.9r_S$. Opposite to Region I there is Region III and opposite to Region II there is Region IV, known as the *white hole solution*, where the time seems to go in reverse. It is clear enough that both are artifacts and have no possible physical meaning

and in Region 2, that is, for $r < r_S$ by

$$u = \sqrt{1 - \frac{r}{r_S}} \exp\left(\frac{r}{2r_s}\right) \sinh\left(\frac{ct}{2r_S}\right) \tag{8.53}$$

$$v = \sqrt{1 - \frac{r}{r_S}} \exp\left(\frac{r}{2r_s}\right) \cosh\left(\frac{ct}{2r_S}\right), \tag{8.54}$$

$$v^2 - u^2 = \left(1 - \frac{r}{r_S}\right) \exp\left(\frac{r}{r_S}\right). \tag{8.55}$$

With this transformation, the metric becomes

$$ds^2 = \frac{4r_S^3}{r} e^{-\frac{r}{r_S}} (du^2 - dv^2) + r^2(d\theta^2 + \sin^2(\theta) d\phi^2), \tag{8.56}$$

where $r(u, v)$ is r expressed as a function of u and v. The angular dependence is $(r^2(d\theta^2 + \sin^2\theta d\phi^2)$, with $r = r(u, v)$. However, if we interpret this solution as representing a *real black hole* that is a collapsing object, there is no possibility for us to explore the inside of the horizon or to receive information from an infalling observer after a finite time; therefore, this solution has no physical meaning. Indeed, each observer must describe the reality on the basis of his own measurements, and this is a key assumption of Relativity. Actually, only Region 1 has a physical meaning as the long time limit of the collapse. Interpretations of regions 3 and 4 in terms of white holes are mere mathematical extrapolations.

8.5.3 *Eternal Black Holes, Daring Hypotheses and Wild Fantasies*

One can verify that the metric (8.56) solves the Einstein equations everywhere outside the point mass (which is often called a singularity, but has been assumed to be there from the start). The diagram is quite effective in demonstrating the properties of the black hole. Suppose one wishes to send a signal from outside to the singularity. The starting space–time point must be chosen in Region 1. The propagation takes no proper time, so $ds = 0$ and $du = \pm dv$. Thus, the word line of the signal will be parallel to the $v = -u$ axis and will hit the singularity, necessarily after crossing the $t = \infty$ line. The physical meaning is that an observer in Region 1 will have to wait an infinite time to see the signal crossing the horizon. Similarly, if a material body is fired to the singularity, it will fall to it and be swallowed in a finite proper time; however, an outside observer will see it accelerating according to the Newtonian physics, but then decelerating, becoming dim and never reaching the horizon. The slowing down of clocks (or redshift phenomenon) eventually wins. The proper time interval needed for a signal to reach an external observer at rest from the horizon diverges and the horizon is a surface of infinite redshift. A signal received by a stationary observer at $r = r_{out}$ from a point at r has a frequency $\omega = \sqrt{\frac{g_{00}(r)}{g_{00}(r_{out})}}$ which vanishes for $r = r_S$.

On the other hand, an observer falling into the black hole would receive an increasingly blue shifted radiation while approaching the horizon and would be fried by this radiation, see all the evolution of the universe before entering the horizon. It would be able to visit the interior after the end of the Universe.

The Galilean method is the key of progress and not an optional. In my opinion, it is clear enough that all the stories about visiting the singularities and then other Universes are not scientific. They cannot be verified or falsified in any way.

Another evident implication of the above is that in all the Universe no actual black hole with horizon and singularity, as described in popular books, exists. The singularity was not discovered but assumed at the start as a simple convenient model by Schwarzschild. The infinite slow-down of the infalling matter approaching a horizon evidently implies that the horizon itself takes an infinite time to complete. The

Penrose theorem that the existence of a horizon implies the existence of a singularity has consequences in the limit $t \to \infty$. The actual black hole is a neverending gravitational collapse. The Schwarzschild black hole model with the singularity and the addition of a nice and done horizon is often called *eternal black hole*.

8.5.4 Shift of the Perihelion

The Kepler problem leads to closed orbits for the reasons discussed in Sect. 2.6, but in the Solar System the perihelion of each planet shifts a little at each orbit. Already in the Newtonian approximation, the many-body problem is formidable because of the mutual interactions, and such systems tend to be unstable in the long run[14]; however, in our solar system the corrections that cumulate in one orbit of each planet are so small that one can simply add the effects of each binary interaction, and at the beginning of the XX century, it was known that the sum of the corrections added to a total 5557 arc seconds per century, while the observed shift of the perihelion was 5600 arc seconds. This was indeed explained by Einstein as another small, additive correction of relativistic origin. One can model the situation as if Mercury was a test particle moving in a geodesic of the Schwarzschild metric (8.47). Then, the nonzero elements of the metric are

$$g_{00} = -\left(1 - \frac{r_S}{r}\right) = \frac{1}{g^{00}}, g_{11} = \left(1 - \frac{r_S}{r}\right) = \frac{1}{g^{11}},$$

$$g_{22} = r^2 = \frac{1}{g^{22}}, g_{33} = r^2 \sin^2(\theta) = \frac{1}{g^{33}}.$$

The motion can be taken to be confined to the plane $\theta = \frac{\pi}{2}$. The geodesic the planet follows can be derived from the Lagrangian (7.26), which, in this metric, is given by

$$\mathcal{L}_{Geo} = \frac{1}{2}\left[\frac{\dot{r}^2}{1 - \frac{r_S}{r}} + r^2\dot{\phi}^2 - \left(1 - \frac{r_S}{r}\right)\left(\frac{d(ct)}{d\tau}\right)^2\right]. \tag{8.57}$$

The proper time of the planet is $\int d\tau$ along the space–time geodesic, which is given by the action principle using the above Lagrangian; therefore, the *action integral* is proportional to the proper time. We define the proper time as given by

[14] The orbits of many body systems are chaotic over long timescales. The Solar System possesses a Lyapunov time, perhaps in the range of a hundred million years (the Lyapunov time is the characteristic timescale on which a dynamical system is chaotic) although we know that life has definitely existed for a longer time than that. However, the relativistic shift has been observed through its cumulative effects on the scale of a century.

$$\frac{\dot{r}^2}{1-\frac{r_S}{r}} + r^2\dot{\phi}^2 - \left(1-\frac{r_S}{r}\right)\left(\frac{d(ct)}{d\tau}\right)^2 = -c^2. \tag{8.58}$$

As in the classical case, we have two conservation laws arising from cyclic coordinates, namely, ϕ and t. The former yields $\frac{\partial L}{\partial \dot{\phi}} = 0$, that is, the conservation of angular momentum

$$r^2\frac{d\phi}{d\tau} = L, \tag{8.59}$$

where L is a constant of integration, essentially the angular momentum of the planet divided by its mass; the latter conservation law is $\frac{\partial L}{\partial \dot{t}} = 0$ and gives us

$$\left(1-\frac{r_S}{r}\right)\dot{t} = Z, \tag{8.60}$$

where Z is another constant. We eliminate the time derivative by writing $\frac{dr}{dt} = \frac{dr}{d\phi}\dot{\phi}$. Putting (8.60) and (8.59) in (8.58), one arrives at

$$c^2\left(1-\frac{r_S}{r}\right)\left(\frac{\dot{t}}{\dot{\phi}}\right)^2 - \frac{\left(\frac{dr}{d\phi}\right)^2}{1-\frac{r_S}{r}} - r^2 = \left(\frac{c}{\dot{\phi}}\right)^2.$$

Now we may set

$$\frac{\dot{t}}{\dot{\phi}} = \frac{dr}{d\phi}$$

and obtain the differential equation

$$(\frac{dr}{d\phi})^2 = \frac{c^2Z^2r^4}{L^2} - r^2\left[1+\left(\frac{cr}{L}\right)^2\right]\left(1-\frac{r_S}{r}\right), \tag{8.61}$$

which determines the angular dependence of the radial coordinate. As long as we can identify r with the distance from the origin (which is correct for weak fields) this can be understood as the equation for the orbit. In order to make the equation slightly simpler, we set $u = \frac{1}{r}$, which implies $\frac{dr}{d\phi} = \frac{-1}{u^2}\frac{du}{d\phi}$. One finds, after multiplying both sides by u^4,

$$\frac{c^2(Z^2-1)}{L^2} + \frac{c^2r_S}{L^2}u + r_Su^3 = u^2 + \left(\frac{du}{d\phi}\right)^2. \tag{8.62}$$

This could be integrated, but it is customary to differentiate with respect to ϕ and get

$$\frac{d^2u}{d\phi^2} + u = A + Bu^2, \tag{8.63}$$

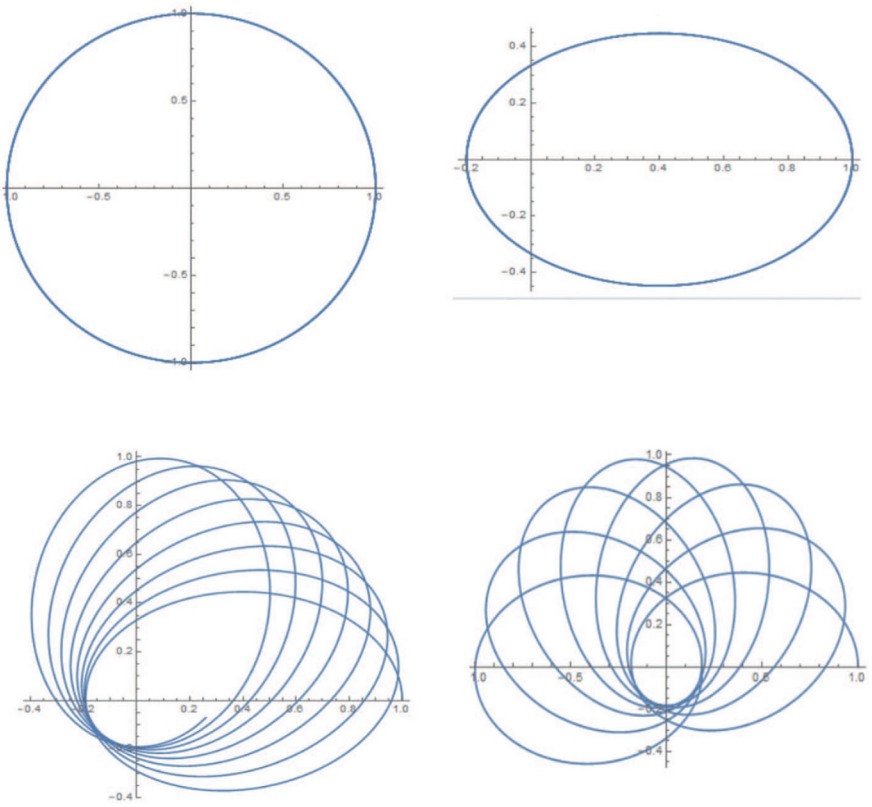


Fig. 8.3 Orbits obtained by a polar plot of the inverse $r = \frac{1}{u}$ of numerical solutions of Eq. (8.62)

where $A = \frac{GMm^2}{L^2} = \frac{c^2 r_S}{2h^2}$, $B = \frac{3GM}{c^2} = \frac{3}{2} r_S$ and $h = \frac{L}{m} = r^2\omega$ is the classical angular momentum of the planet divided by its rest mass; this is the generalization of Eq. (2.41) of Classical Mechanics, that in the present notation becomes $\frac{d^2u}{d\phi^2} + u - \frac{GMm^2}{L^2} = 0$. Even in the full theory, the mass of the planet does not enter into it. Equation (8.63) can be studied numerically on a laptop, see Fig. 8.3. It turns out that for $B = 0$, we get the classical closed orbits which are circular if $A = 1$. The extra term B produces a rosetta-like orbit and the deviations from the Newtonian theory are large in the case of compact binary stars.

The great accuracy of the Newton theory makes it clear that B is very small, and so one can trivially solve the problem through perturbation theory, i.e., by approximating the factor of B by the Newtonian solution. One easily finds that the rate of precession of the perihelion is

$$\frac{\Delta\phi}{T} = \frac{3\pi r_S}{Ta(1-e^2)}, \tag{8.64}$$

where a is the semi-major axis of the ellipse and e its eccentricity. The two are bound by the relation $a(1-e^2) = \frac{L}{GMm^2}$, where L is the angular momentum. The calculated shifts are 42.9" per century for Mercury, 8.6" for Venus, 3.8" for the Earth and 1.35" for Mars. They have been confirmed by the observations; the main residual uncertainties are due to small complications like the quadrupole moment of the sun.

8.5.5 Geodetic Effect

Consider a gyroscope on a satellite in a circular orbit around a mass, e.g., the Earth. An observer on the satellite is in free-fall, and therefore notices that the gravity field is locally removed; he establishes a frame of reference with the origin on the satellite and axes oriented in such a way that the fixed stars do not move. Finally, he sets a gyroscope in a fixed position with respect to his reference. According to the Newtonian mechanics, both the observer on the satellite and an observer on Earth agree that no force and no torque are acting on the gyroscope which should keep rotating around the Earth in the same way forever, keeping its axis forever in the same direction. In General Relativity, there are three distinct effects that cause tiny changes of the rotation axis at each rotation around the Earth. Two effects are related to the rotation of the Earth, and will be described briefly in the next section. The present section is devoted to the geodetic effect (sometimes called the geodesic effect), which occurs because the satellite is rotating around a mass, independently of the rotation of the latter. The mass produces a metric close to the Schwarzschild metric in the reference of the Earth; this metric does not mix time with the angles and there is no precession for gyroscopes on Earth. However, the instantaneous frame on the satellite is connected to the Earth frame by a Lorentz transformation, and this does mix space and time. Therefore, the metric tensor on the satellite depends on time; this produces a time dependence of the components of the spin in the satellite reference. For symmetry reasons, a gyroscope which is set in the plane of the orbit must remain in the plane. The rate of change of the spin $\boldsymbol{s}$ of a gyroscope in orbit with radius r and velocity v due to this effect is found, using a weak field approximation, to be

$$\frac{d\boldsymbol{s}}{dt} = \boldsymbol{\Omega}_G \wedge \boldsymbol{s}, \tag{8.65}$$

where

$$\Omega_G = \frac{3}{2}\frac{r_S}{r^3}\boldsymbol{r} \wedge \boldsymbol{v}. \tag{8.66}$$

If the gyroscope points away from Earth and turns anticlockwise, the spin is positive. Then, for each rotation, the axis gains an angle of order $\alpha = \frac{3}{2}\pi\frac{r_S}{r}$. The effect was verified experimentally in 2007 with an accuracy better than 0.5 per cent by the

Gravity Probe B cooperation,[15] using a satellite in a polar orbit at 642 km from Earth; the shift was in the direction of motion and was as small as 0.0018 °C per year.

8.5.6 *Relativistic Delay of Signals in a Gravity Field*

Irwin Shapiro, in 1964, suggested a test of General Relativity based on the following principle. A planet between Earth and the Sun or beyond the sun but visible near the solar disk can be used as a mirror that reflects a radio signal. The signal emitted from Earth at r_E and reflected at the planet in r_R will explore the metric generated by the solar mass, approximately given by the Schwarzschild expression (8.47). The coordinate time length of the outward voyage can be obtained putting $d\tau = 0$ and $d\theta = d\phi = 0$ (since the signal travels at c) in (8.47). Integrating, and multiplying by 2 to include the return trip, one finds that

$$\Delta t = \frac{2}{c}\int_{r_E}^{r_R} \frac{dr}{1 - \frac{r_S}{r}}.$$

To convert it into the proper time, we use Sect. 8.6, do the integral and find

$$\delta\tau = \frac{2}{c}\sqrt{1 - \frac{r_R}{r_E}}\left[r_R - r_E + r_S \ln\left|\frac{r^E - r_S}{r^R - r_S}\right|\right]. \tag{8.67}$$

This is the coordinate time interval. The Shapiro effect consists in the fact that this is not the length of the trip divided by c. Indeed, the gravity field of the sun causes a delay.

Measurements performed on upper conjunctions of Venus (i.e., Venus close to the sun in the sky but beyond the sun) were already performed in 1970 and more measurements were done by the rockets Mariner 6 and Mariner 7, giving excellent agreement with the theory.

One can be surprised by the apparent reduction of the speed of light from c to something like $c(1 - \frac{r_S}{r})$, where r is the distance from the sun. It appears to contradict the principle of equivalence. However, this is an effect of curved space–time; it can be rationalized in terms of an increase of the radial distances and a slowing down of the time in the neighborhood of important masses. But any local measurement done by taking rods and clocks into the space close to the sun would give c for the speed of light, because rods and clocks would be affected themselves by the field.

[15] See Everitt, et al. (2015).

8.6 Clocks in a Gravity Field, GPS and Gravitational Redshift

When discussing the Schwarzschild solution, we have already noted the effect of gravity on the proper time τ, which is evident in Eq. (8.48), namely, $dt = \frac{d\tau}{\sqrt{1-\frac{r_S}{r}}}$; here $t > \tau$ is the coordinate time measured by an observer at infinity and τ is the time measured by at distance r from the origin. For weak fields, when the Newton theory is applicable, the (positive) gravitational potential is

$$\Phi(r) = \frac{MG}{r} \tag{8.68}$$

and one can write

$$dt = \frac{d\tau}{\sqrt{1 - 2\frac{\Phi(r)}{c^2}}}.$$

Let us perform some thought experiments in the gravity field described by the Schwarzschild metric (8.47) of a massive body. The first experiment consists in the emission of a light signal from a point (r_E, θ_E, ϕ_E) at time t_E and its detection at the space–time point $(r_R, \theta_R, \phi_R, t_R)$.

The light follows a null geodesic in space–time, and the time it takes, i.e., $t_R - t_E$, depends only on the two space points and can be obtained from the metric once the space trajectory of the light is specified; finding the trajectory is a problem of geometry that can be solved once and for all and does not concern us in this example. If the experiment is repeated after a delay Δt_E, the same delay affects the reception, so $\Delta t_R = \Delta t_E$. This entire story looks the same as that of the Newton dynamics, but it is different, because the local clocks measure the proper time τ, not t. This implies that

$$\Delta\tau_R = \Delta\tau_E \sqrt{\frac{1 - \frac{r_S}{r_R}}{1 - \frac{r_S}{r_E}}}. \tag{8.69}$$

The actual repetition frequency is modified by the gravity field.

In the next thought experiment, instead of measuring the repetition frequency we measure the color of the light and say that its frequency depends on the position in the field of the source. The ratio of frequencies is

$$\frac{\nu_R}{\nu_E} = \sqrt{\frac{1 - \frac{r_S}{r_E}}{1 - \frac{r_S}{r_R}}}. \tag{8.70}$$

The weak field case can be obtained by expanding this result for small r_S; the same result is obtained setting $ds = cd\tau$ and $d\boldsymbol{r} = 0$ in the approximate metric (8.44).

The proper time τ of a clock is related to the time t measured by an identical clock outside the field by

$$d\tau = \sqrt{\left(1 - \frac{2\Phi}{c^2}\right)}dt \sim \left(1 - \frac{\Phi(r)}{c^2}\right) dt. \tag{8.71}$$

Since $U = 0$ at infinity and $U < 0$ in the field, a distant observer will conclude that the field makes the clocks run slow. Of course, the variable r that appears in these expressions is a coordinate radius, which is not simply the distance from the origin in the case of strong fields, near the horizon.

The gravitational redshift of spectral lines in the light coming from the sun, and particularly from dense stars like white dwarfs, was observed long ago.

For an atom A that emits at a proper frequency ν_A at distance r_A from a mass M, an observer B at distance r_B from M will see light at frequency ν_B

$$\frac{\nu_B - \nu_A}{\nu_A} \sim \frac{GM}{c^2}\left(\frac{1}{r_B} - \frac{1}{r_A}\right), \tag{8.72}$$

assuming that the fields are weak (the gravitational energy must be small compared to the rest energy). For an atom in the gravitational field of a star with radius R and Schwarzschild radius r_S, this implies a redshift of the frequency given by

$$\frac{\Delta\nu}{\nu} = \frac{r_S}{2R}. \tag{8.73}$$

For the Earth, $M = 6 \times 10^{24}$ kg, $r_S = 9 \times 10^{-3}$ m and $\frac{\Delta\nu}{\nu}$ is on the order of 10^{-9}; for the sun, $M = 2 \times 10^{30}$ kg, $R = 7 \times 10^8$ m, $r_S = 3$ km and the shift is still one part in a million; but for a neutron star with one solar mass and a radius of order 10^4 m one expects $\frac{\Delta\nu}{\nu}$ is in the scale of 10^{-1}.

A measurement of the gravitational redshift in the laboratory was performed for the first time in 1965 by Pound and Snider using the Mössbauer effect, a laser and the Earth's gravity. Now this is routine. The clocks on the satellites used for the Geo-positioning System must be corrected to account for relativistic effects. They go slow (compared to clocks on Earth) by 7 ms/d because of the time-dilation effect of Special Relativity, but they go fast by 45 ms/d because of the gravitational effect. Such small times are important in practice, because light is so fast. Without the resulting correction of 38 ms/d the GPS of our cars would have errors on the order of kilometers.

8.7 Bending and Gravitational Lensing of Light

The deflection of starlight by the gravity of the Sun was one of the first tests of General Relativity during a solar eclipse in 1919. Let us put the sun at the origin of coordinates and let the ray come parallel to the x axis, with the impact parameter R. We take the wave four-vector k with $k_0 = \frac{\omega}{c}$ and $\overrightarrow{k}(x)$ such that $\overrightarrow{k}(x = -\infty) = (k, 0, 0)$. We must determine how k_y grows with x.

Consider the propagation of a ray according to Eq. (8.30),

$$\frac{dk^\mu}{dx^\beta} + \Gamma^\mu_{\alpha\beta} k^\alpha = 0;$$

since we suppose that the ray arrives along the x axis ($\beta = 1$) with wave vector parallel to the x axis ($\alpha = 1$) and is deflected in the y direction ($\mu = 2$), we need the Christoffel Γ^2_{11}.

This connection is due to the weak field produced by a mass M centered at the origin, whose approximate metric, in terms of the Newtonian gravitational potential V, is given by (8.44), namely,

$$ds^2 = -\left(1 + \frac{2V}{c^2}\right) c^2 dt^2 + \left(1 - \frac{2V}{c^2}\right) d\boldsymbol{r}^2.$$

We need nothing more than the lowest-order effects.

Assume that the ray is deflected by a small angle, which can be approximated by $\alpha = \frac{k^2(\infty)}{k^1}$.

Since the covariant components of g are

$$diag(-(1 + \frac{2V}{c^2}), (1 - \frac{2V}{c^2}), (1 - \frac{2V}{c^2}), (1 - \frac{2V}{c^2})),$$

the contravariant components are immediately found using (7.11) and we readily obtain the affine connection directly from Eq. (7.27), namely,

$$\Gamma^k_{ij} = \frac{1}{2} g^{im} \left(\frac{\partial g_{jm}}{\partial x^k} + \frac{\partial g_{mk}}{\partial x^j} - \frac{\partial g_{kj}}{\partial x^m} \right).$$

One finds $\Gamma^2_{11} = \frac{1}{c^2} \frac{\partial V}{\partial y}$. Therefore,

$$\frac{dk^2}{dx} = \frac{-2\omega}{c^3} \frac{\partial V}{\partial y}. \tag{8.74}$$

Integrating, one finds the angle

$$\alpha = -\frac{4GM}{Rc^2} = -\frac{r_S}{R}, \tag{8.75}$$

where r_S is the Schwarzschild radius. The massless light is deflected like a particle of mass M that feels an acceleration $g = \frac{MG}{R^2}$ along a path length $2R$; indeed, setting $x \sim ct$, $y \sim \frac{gt^2}{2}$, one gets $y'(x) \sim \frac{GMx}{R^2c^2}$, and for $x = 2R$, $\tan(\alpha) \sim \alpha$ yields Eq. (8.75).

In the case of the sun, inserting $r_S = 3\,\text{km}$ and $\text{R} = 7 \times 10^5$ km in Eq. (8.75) one finds 8.57×10^{-6} which corresponds to 1.75 arc s. This was first verified approximately in an eclipse in 1919 and has been so repeatedly with increasing precision since then.

The empty space behaves like a non-dispersive dielectric medium because of the distortion due to gravity. This is very important in astrophysics. Galaxies are often seen to distort and amplify the images of background objects, often showing arcs, so-called Einstein rings or multiple images of the same object. In Sect. 9.5, we shall see that single photons may form multiple images corresponding to different routes from a Quasar around a Galaxy to Earth, and then interfere with themselves. This is a very large-scale test of Quantum Mechanics!

8.8 Kerr Black Hole, Frame Dragging and Gravitomagnetic Field

Since all forms of energy, including kinetic energy, contribute to the gravity, one must expect that in the relativistic theory, the rotation of the sun must have some influence on the motion of the planets. This is true, but there is more. If the mass rotates, with angular momentum J, the metric (8.47) is generalized by the Kerr metric, discovered by Roy Kerr in 1963, almost 50 years after the Schwarzschild solution, since it is quite involved. This was an extraordinary tour de force, achieved when no symbolic manipulation software existed. It is much more realistic than the Schwarzschild solution, since all stars rotate, but it is also more problematic, since it is still a long time limit of a collapse, for the same reasons as in the Schwarzschild case, with a similar horizon. However it is still a vacuum solution, and since a point mass cannot provide any angular momentum, this must be provided by the field; instead in any collapsing body the matter also contributes. This unphysical feature probably produces a ring-like singularity and solutions indicating closed time paths; for such pathologies, the reader should consult the advanced literature.[16] Anyhow, it is clear that they are artifacts of the infinite time vacuum solution and cannot be taken as real effects. Instead, we shall concentrate on some results that have observable consequences. Besides the Schwarzschild radius r_S, a new length enters, namely,

$$\alpha = \frac{J}{Mc}, \tag{8.76}$$

[16] See for example Matt Visser (2008).

where M is the mass; moreover, we need

$$\Delta(r) = r^2 - rr_s + \alpha^2. \tag{8.77}$$

In terms of the usual spherical coordinates, the Kerr metric is

$$ds^2 = (1 - \frac{r_S r}{\rho^2})c^2dt^2 - \frac{\rho^2}{\Delta}dr^2 - \rho^2 d\theta^2 - (r^2 + \alpha^2 + \frac{r_S r\alpha^2}{\rho^2}\sin^2(\theta))\sin^2(\theta)d\phi^2 + \frac{2r_S r\alpha\sin^2(\theta)}{\rho^2}cdtd\phi. \tag{8.78}$$

This solution can be interpreted as a rotating *eternal black hole*.

The Kerr metric has the direction of the angular momentum as a preferred direction; the singularity is surrounded by an event horizon of radius $r = M + \sqrt{M^2 - \alpha^2}$ and by an *ergosphere* delimited by a *limit of stationariness* within a surface of equation $r\theta = M + \sqrt{M^2 - \alpha^2\cos^2\theta}$, ($\theta = 0$ in the direction of the angular momentum) (Fig. 8.4).

The ergosphere is a region of extreme *frame dragging*, so it is time to introduce this phenomenon.

Josef Lense and Hans Thirring predicted a rotation of the orbit of a test particle around a rotating body already in 1918 and applied their findings to the solar system. The Lense–Thirring effect depending on the angular momentum J of the Sun consists of a precession of the nodal plane, which is the intersection of the orbital plane of a plane with the equatorial plane of the Sun; this is independent of the perihelion precession discussed above, which does not depend on the rotation of the Sun. It

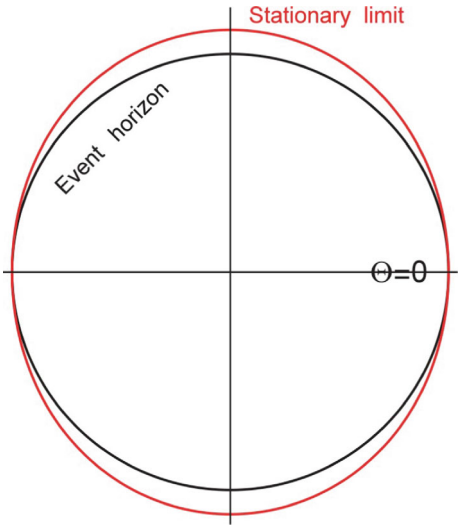


Fig. 8.4 The singularity in the Kerr solution is surrounded by an event horizon (black) and by a "limit of stationariness" (red); here M = 1 and a = 0.6; the angular momentum is in the horizontal direction

depends on the semi-major axis a of the orbit on its eccentricity e. The angular frequency of rotation is $\frac{2J}{a^3(1-e^2)^{\frac{3}{2}}}$.

The Lense–Thirring rotation on the orbit of a satellite around the Earth is extremely slow, but the effect has been verified by the LAGEOS satellite (Laser Geodynamics Satellite) in 2004, and with better accuracy by the Gravity Probe B cooperation more recently.

The most characteristic term in the Kerr metric is the one in $d\phi dt$. The reader might have noticed that a mixing of time and space in the metric occurs here and in the rotating platform problem that we studied at the beginning of this chapter. The Coriolis inertial force has a gravitational analogue, namely, gravito-magnetism. Actually, in the weak field case, $\boldsymbol{A} = (h_{01}, h_{02}, h_{03})$ is like a three-vector in analogy with the electromagnetic vector potential; its curl H is the gravito-magnetic field $\boldsymbol{H}$. This produces a *gravitational Lorentz force*

$$\boldsymbol{F} = m\boldsymbol{v} \wedge \boldsymbol{H}, \tag{8.79}$$

which acts on the masses in a way that resembles the magnetic force acting on charges. The force (8.79) also produces a torque on gyroscopes. The effect is similar to the precession of a magnetic moment in a magnetic field. If we put the origin at a massive body with angular momentum $\boldsymbol{J}$, the angular velocity of precession of a gyroscope at $\boldsymbol{x}$ can be shown to be $-\frac{H}{2}$.

A falling gyroscope, which is not rotating initially, defines an instantaneous inertial system, like any test particle. The remarkable effect is that if the field is produced by a rotating body, the instantaneous inertial system is rotating, while a non-rotating system would not be inertial. Therefore, one speaks about a *dragging of inertial frames*. The rotation of the Earth causes the rotation of the inertial system in the satellites. The effects on artificial satellites are very small and tend to be masked by other effects, like irregularities in the Earth's field; however, the predictions of General Relativity have been tested and confirmed by accounting accurately for all the disturbing effects. According to predictions by Pugh and Schiff, this frame dragging effect produces a drift rate of the angular momentum $\boldsymbol{s}$ for a gyroscope on a satellite at distance r from the center of the Earth[17]

$$\frac{d\boldsymbol{s}}{dt} = \boldsymbol{\Omega}_{fd} \wedge \boldsymbol{s} \tag{8.80}$$

with the frequency

$$\boldsymbol{\Omega}_{fd} = \frac{GI}{c^2 r^3}\left[\frac{3\boldsymbol{r}}{r^3}(\boldsymbol{\omega}_e \cdot \boldsymbol{r}) - \boldsymbol{\omega}_e\right], \tag{8.81}$$

where I is the moment of inertia and $\boldsymbol{\omega}_e$ the angular velocity of our planet.

In the geometry chosen by Gravity Probe B, as mentioned above, the orbit is polar and the frame dragging effect 1.1×10^{-5} degrees per year is well-separated from the

[17] See Everitt, et al. (2015).

Geodetic effect, because the two shifts are at right angles, with the frame dragging producing a shift toward the West while the Geodetic effect acts in the North–South direction. In the ergosphere of a rotating black hole, the frame dragging is so strong that all masses are forced to rotate around the black hole; however, in principle, an object can enter it and leave. Roger Penrose suggested that masses could be thrown in and extracted with increased energy, obtained at the expense of the black hole rotation energy.

8.9 Gravitational Waves

In Astrophysics one applies the results of experiments done in laboratories on Earth and surroundings to the study of stars, galaxies and even larger bodies. General Relativity and in particular the gravitational waves are very important tools in this field. It is time to resume, from Sect. 8.4.1, the analogy and the differences with electromagnetism in the production of radiation. From Eq. (8.43), the linearized equations for the correction to $\eta_{\mu\nu}$ in vacuo are just the familiar wave equations

$$\Box h_{\mu\nu} = 0; \tag{8.82}$$

for waves propagating along the z axis, the solutions are of the form

$$h_{\mu\nu} = h_{\mu\nu}(kz - \omega t + \phi_{\mu\nu}), \tag{8.83}$$

where $k = \frac{\omega}{c}$ and $\phi_{\mu\nu}$ are constant phases. Using these, one must look for the metric tensor. One can show that there are two polarizations. The general form of the metric associated with the plane waves propagating along z is of the unsymmetric form

$$ds^2 = (1 + h_{xx})dx^2 + 2h_{xy}dxdy + (1 - h_{xx})dy^2 + dz^2 - c^2dt^2; \tag{8.84}$$

there are solutions of the xx type (with $h_{xy} = 0$) and solutions of the xy type (with $h_{xx} = 0$). The first kind tends to stretch and compress objects along the x and y axes in phase, while the second does the same at 45 °C.

Since there are no negative masses, there are no dipoles and there is no dipole radiation. The leading contribution is quadrupolar. From the field equations, one can derive the energy P radiated per unit time when masses are accelerated.

$$P = \frac{G}{5c^5} \sum_{\alpha,\beta=1}^{3} (\frac{d^3 Q_{\alpha\beta}}{dt^3})^2, \tag{8.85}$$

where

$$Q_{\alpha\beta}(t) = \int \rho(x,t)(x_\alpha x_\beta - \delta_{\alpha\beta}\frac{r^2}{3})d^3x \tag{8.86}$$

are the components of the quadrupole tensor. Here is the Fourier transform of the formula describing the power radiation of electric quadrupole electromagnetic waves.[18]

$$P = \frac{ck^6}{360} \sum_{\alpha,\beta} |Q_{\alpha\beta}|^2. \tag{8.87}$$

Apart from the factor in front of the summation, the similarity is evident. For a periodic motion, Eq. (8.85) implies that $P \sim \frac{G\omega^6}{c^5} \sum_{\alpha,\beta=1}^{3} Q_{\alpha\beta}^2$.

8.9.1 Astrophysical Sources

The binary pulsar PSRB1913 + 16 discovered in 1974 by Russell Alan Hulse and Joseph Hooton Taylor provided an indirect but compelling confirmation of gravitational waves; for this discovery, they received the Nobel prize in 1993. It is a system of two neutron stars, one of which is a pulsar. Pulsar emits a narrow cone of radiation while rotating very fast, and the radiation arrives on Earth as a series of regular pulses. In this case, the period is about 59 ms. The two neutron stars are close to each other and rotate around their center of gravity in just about 7.75 h. The pair emits energy in the form of gravitational energy, and this causes a measurable decrease of the period (and eventually, the two stars will collide and coalesce, producing a strong blast). The observations were in agreement with the calculations based on General Relativity.

Finally, gravitational waves were detected directly for the first time in September 2015 by researchers of Caltec, MIT and LIGO; the data were interpreted as being due to the merge of two black holes that occurred 1,3 billion years ago. The signal frequency increased from 35 to 250 Hz during the last orbits of both objects approaching merge. The amplitude of the oscillations was in the range of 10^{-15} m. It was possible to determine the masses (29 and 36 solar masses) and it was estimated that about 3 solar masses were emitted in the form of gravitational waves. Several similar events have been reported since then, even in coincidence with the Virgo gravity wave observatory in Cascina (Pisa, Italy). For this enormous achievement, the Nobel Prize for Physics for 2017 has been awarded to the U.S. scientists Kip Thorne, Barry Barish and Rainer Weiss and the international cooperations Ligo and Virgo have been mentioned in the motivation. The detection of gravity waves was hailed as the beginning of a new branch of Astronomy. Indeed, in 2017, the LIGO and Virgo collaborations detected gravitational waves originating from the coalescence of a binary neutron star system, which also originated a gamma-ray burst observed by the Fermi gamma-ray Space Telescope; the finding was confirmed by optical telescope observations of the same event. As we shall see in the next paragraph, the

[18] See, e.g., John D. Jackson, "Classical Electrodynamics", Sect. 9.3.

cosmic background radiation is the most ancient object that can be seen by electromagnetic radiation. By gravitational waves, instead, we can hope to gain information on the evolution of the universe during the first phases of its expansion.

8.9.2 Gravitons?

Some theoreticians are dissatisfied with General Relativity since it is classical; they are sure that it must be *quantized*, despite the fact that gravity is fundamentally different from the other interactions (it is not a gauge theory). This topic is too technical to be discussed in a book like this; besides being technical, those models are also inconclusive, since there are no results that can be compared to experiment. However, all approaches to quantum gravity (strings, loop quantum gravity, noncommutative geometry) pretend that gravitational waves must be quantized, and their tensor character demands that the quantized waves (or gravitons) must be spin 2 Bosons, in the same way as vector electromagnetic waves have spin one. However, these theories start with arbitrary conjectures about some structure of space–time at the Planck level, which is many orders of magnitude away from anything we know. In order to give a sensible physical meaning to a theory of gravitons, one must conceive a working graviton detector and graviton interactions like a gravitational Compton effect. Otherwise, we do not know what we are speaking about.

It is remarkable that the great theoretician Freeman Dyson has questioned the possibility[19] of detecting a graviton, assuming that such a particle exists, because the cross section would be so small that a Jupiter-sized detector would be needed. This concept has been expanded.[20] This is a radical criticism of the opinion that gravity must be quantized. Never mind, we can keep it (and space–time) classical until proven otherwise.

8.10 The Standard Model of Cosmology and the Big Bang

Since Galileo pointed his first telescope at the sky and modern astronomy was born, a great quantity and variety of phenomena have allowed Astrophysics to develop. Now we have credible models of stars, black holes and galaxies. In addition, since the discovery of the expansion of the Universe by Edwin Hubble the ancient idea of an ethernal immutable Universe was trashed. Cosmology is extremely mind-bending application of Theoretical Physics to the Universe and its origin[21] but the starting

[19] Dyson (2004).

[20] Tony Rothman and Stephen Boughn (2006).

[21] On the other hand, the deeper it pushes into the past, the further one moves away from the conditions that can be tested, the greater the risk of accepting unfounded and even completely unscientific theories and keeping them as true. We risk to replace the ancient myths by other myths.

point that embodies the simplest consequences of General Relativity in this field is rather simple. The Newtonian theory cannot describe properly an infinite Universe (but only a very large finite one) and in the case of a finite Universe, it cannot treat all points on the same footing, as required by the Copernican Principle (no provileged place). The relativistic standard model is based precisely on these symmetry properties, namely, isotropy and homogeneity, supposed to hold on very large scale much above galaxy clusters. It is this homogeneity which allows for a simple solution of the Einstein equations which would otherwise be hard to handle. The stress–energy tensor will be independent of position; moreover, we can choose a co-moving reference, that is, we may set the velocity of matter to zero in every point; then, the stress–energy tensor (6.30) has only one nonzero component, namely, $T_0^0 = -(\rho c^2 + \rho U)$, and this component does not depend on position. The component g_{00} of the metric tensor determines the proper time of a body at rest, so it cannot depend on position. We know that the components $g_{0k}, k = 1, 2, 3$ have to do with rotations, and the above principles strongly suggest that we can safely take $g_{0k} = 0$. In conclusion, we are led to a metric of the same simple form as in Special Relativity,

$$ds^2 = -dl^2 + c^2 dt^2$$

and the three-dimensional space metric $g_{ik}, i, k = 1, 2, 3$ yielding dl^2, must be such that the curvature is the same throughout. We have already met such constant-curvature three-dimensional spaces in Sect. 7.2 and we know that basically we have the choice between a closed Universe with space metric proportional to (7.36) and an open Universe (7.38). The flat, Euclidean Universe is the limiting case between the two and the data do not allow to choose; so the Universe may be finite, but very very large, or infinite.

8.10.1 The Relativistic Models

The closed Universe arises from the space metric (7.36) yielding the space–time metric:

$$ds^2 = c^2 d\tau^2 - a^2[d\chi^2 + \sin^2(\chi)(d\theta^2 + \sin^2(\theta)d\phi^2]. \tag{8.88}$$

This becomes flat for $a \to \infty$ (recall the definition $\sin(\chi) = \frac{r}{a}$).

Similarly, the open Universe arising from Eq. (7.38) has the space–time metric

$$ds^2 = c^2 d\tau^2 - a^2[d\chi^2 + \sinh^2(\chi)(d\theta^2 + \sin^2(\theta)d\phi^2]. \tag{8.89}$$

In any case, these four-dimensional models offer a sensible geometrical explanation of the question of a possible beginning of time in a hot Big Bang. The question of the beginning of time worried philosophers as smart as Immanuel Kant, who could not find a solution. But Relativity gives a possible answer: one cannot go back in time

before the Big Bang for the same reason why walking on Earth and starting at the North Pole one can only go south. Alexander Friedmann, Georges Lemaître, Howard P. Robertson and Arthur Geoffrey Walker formulated this model before 1930, without the cosmological term Λ which was not supported by the data available then.

The Einstein equations for these models can be solved exactly; this can be done directly as shown in Landau–Lifshitz[22] and other classics, but the argument is still rather long.

In the case of (8.88) it is expedient to replace the time τ with a new variable η, by putting

$$a = c\frac{dt}{d\eta}; \tag{8.90}$$

then,

$$ds^2 = a^2[-d\eta^2 + d\chi^2 + \sin^2(\chi)(d\theta^2 + \sin^2(\theta)d\phi^2]. \tag{8.91}$$

Now,

$$g_{\mu\nu} = \text{diag}(g_{00}, g_{11}, g_{22}, g_{33}) = \text{diag}(-a^2, a^2, a^2\sin^2(\chi), a^2\sin^2(\chi)\sin^2(\theta))$$

and

$$g^{\mu\nu} = \text{diag}(g^{00}, g^{11}, g^{22}, g^{33})$$

with

$$g^{\mu\mu} = \frac{1}{g_{\mu\mu}}.$$

The result depends on the total mass M of the Universe, which is a conserved quantity. We need the constant $a_0 = \frac{2kM}{3\pi c^2}$, where $k = c^4 G_N$. Eventually, one finds the parametric law governing the size of the Universe a, namely,

$$\begin{aligned} a(\eta) &= a_0(1 - \cos(\eta)), \\ t &= \frac{a_0}{c}(\eta - \sin(\eta)). \end{aligned} \tag{8.92}$$

It turns out that the Universe reaches a maximum size, then collapses to a point and restarts expanding. It is a pulsating Universe with period $T = \frac{2\pi a_0}{c}$. In the case (8.89) the Universe is open and expands forever:

$$\begin{aligned} a(\eta) &= a_0(\cosh(\eta) - 1), \\ t &= -\frac{a_0}{c}(\eta - \sinh(\eta)). \end{aligned} \tag{8.93}$$

[22] L. Landau et E. Lifshitz, Theorie du Champ, editions mir, Moscou 1966, Chapitre XII.

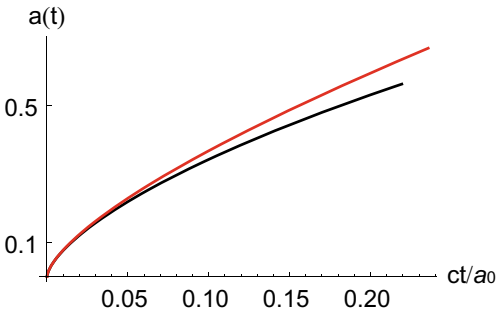


Fig. 8.5 $a(t)$ for the open Universe (red) and for the closed Universe (black) versus time. The red curve eventually increases to infinity linearly, while the black one reaches a maximum and then goes to zero symmetrically, leading to a pulsating Universe. Our Universe is still so young that the two curves (and the flat Universe) are hard to distinguish

It would be exciting to discover that we are in a curved Universe, possibly with a nontrivial topology; the data, however, say that Nature does not like such complications, and the Universe is accurately flat. For a flat Universe,

$$ds^2 = -c^2\tau^2 + b^2(dx^2 + dy^2 + dz^2), \tag{8.94}$$

and one can show that

$$\left(\frac{db}{dt}\right)^2 = \frac{8\pi G}{3c^2}\rho(U + c^2)b^2, \tag{8.95}$$

where it is reasonable to assume that $\rho(U + c^2)b^3$ is constant; then, b grows as a power of t (Fig. 8.5).

Let me pause to stress that the possibility of an infinite Universe has some breathtaking implications. Max Tegmark published an article on Scientific American in 2004 entitled "Parallel Universes". Actually, his argument does not involve separate realities; it involves simply the fact that in an infinite Universe everything that is possible becomes real, no matter how improbable it is; therefore somewhere there are other regions of space that are close to our solar system. By going far enough, one can find situations like ours, identical in every respect. Those regions are considered by Tegmark, arbitrarily, as a type of "parallel universe". He estimates that a twin galaxy exists within (hold fast) $10^{10^{28}}$ m from here. Let me comment that Tegmark's multiverse is more innocent that others. The implications of an infinite universe on the possibility of intelligent life are evident, but what could happen at $10^{10^{28}}$ m from here is not science. However, still more unscientific multiverses pollute the current literature and never produced scientific results.

On the other hand in the case of the finite universe which must have been small at the beginning when the cosmic background radiation was emitted we must admit a superluminal expansion, because we are receiving now the radiation emitted then.

The possibility[23] that the Universe started with a big bang is quite interesting by itself. If the Universe started with a big bang, in the geometry of General Relativity we have a very intriguing, geometric explanation of the beginning of time, a concept that worried Immanuel Kant, who could not explain it. In the Friedmann model, asking what existed before the big bang is as meaningless as the question: what is north of the north pole?

8.10.2 Cosmic Expansion and Background Radiation

Einstein started to apply his theory to the large-scale structure of the Universe in 1917, before the two major discoveries by Edwin Hubble: in 1925 he established the existence of Galaxies outside the Milky Way, leading to a much larger Universe, and in 1929, he discovered the expansion of the Universe.[24] The expansion causes a redshift of the light of a galaxy at distance d; the redshift is like a Doppler shift due to a velocity v given by

$$v = Hd \tag{8.96}$$

where the Hubble constant H is on the order of 45–90 km per second per Megaparsec and 1 parsec is about 3.26 light years. Each Galaxy also has a peculiar velocity, similar to the velocity of molecules in a gas; therefore v is a sort of average value. There is some vague analogy between the galaxies in the expanding Universe and cue balls painted on the surface of an inflating balloon. However, the cue balls on the balloon should grow in size, while the sizes of atoms, the Earth, the Solar System and the Galaxy do not grow: the expansion creates space between bound systems, but not inside them, because they are relatively small and the small effects of the expansion are dominated by the effects of Gravity. Even galaxy clusters remain bound by gravity. The discovery came unexpectedly, but it finally explained the old Olber's paradox. The astronomer Heinrich Wilhelm Olbers (1758–1840) remarked that if the Universe were infinite and static, there should be stars in every solid angle. Indeed the night sky is dark because the farthest stars are redshifted by the Doppler effect.

In 1964, Arno Penzias and Robert Wilson made the fortuitous but very important discovery of the cosmic background radiation at 2.7 K, which confirmed the hot Big Bang, and at the time there was a very optimistic feeling that the concepts and the laws of Physics developed on this Planet were ruling the Universe and although largely incomplete the mosaic was under way. Nobody could question that Cosmology was a branch of Physics. In the 60 s, the exploration of the Solar System started. The relevant distances and masses had been estimated quite accurately by the available telescopes, the parallax and the knowledge of the Newton theory of gravity. The parallax of

[23] Somebody, like the Nobel prize winner Penrose, thinks about a rebound, but this requires new Physics since nobody knows how the entropy could be re-initialized.

[24] In some way Theory and Experiment were in competition since Georges Lemaître had predicted the redshift–distance relation in 1927.

nearby stars is measurable, and since we know the size of the orbit of the Earth the distance can be measured. However, most stellar distances can be measured by more indirect means, and one must rely on the luminosity of peculiar stars (novae, Cefeid variable stars) that can be used as standard candles. This works for the near Galaxies, but the uncertainties grow when dealing with large intergalactic distances, on the order of many millions of light years. At still larger distances, one can rely on the *luminosity distance* $D_L = \sqrt{\frac{L}{2\pi l}}$, where L is the absolute luminosity and l is the power received from the source; another possibility is based on the angular size and the real size of the object. Both methods are affected by the possible evolution of the source in the course of millions or billions of years, and at any rate, the estimate must be based on some assumption on the metric of space–time. At large cosmic distances (many millions or several billions of light years), the relativistic effects are large, as is evident from the large redshifts due to the expansion of the Universe. In addition, it should be kept in mind that in General Relativity, the distance between two objects is meaningful only locally. The distance l loses the obvious meaning of everyday life unless the metric tensor is time-independent; this assumption cannot hold in an expanding Universe. In general one can define the element of distance dl and then compute $\int dl$ over a curved space–time, but the result depends on the path of integration. At any rate, the redshift makes distant objects dimmer and dimmer.

In Special Relativity, no massive body can go faster than light. However, Special Relativity ignores the effects of cosmological expansion; there is no such limitation to the creation of space due to the expansion of the Universe at cosmological distances and therefore no speed limit. It is clear that the major part of the Universe recedes from us at superluminal speeds; on the other hand, we can observe only the objects which recede with speed less than c. It is quite reasonable to believe that the invisible part of the Universe would be familiar to us if we could observe it, but since it is in fact unobservable, whatever happens there is not of scientific interest. With such caveats, the age of the Universe is 13.8 billions of years, and taking the expansion into account, the edge of the observable Universe is estimated to be at 46 billion light years.

8.10.3 Dark Matter

The task of building a coherent scheme in which the observation could be rationalized appeared more feasible 40 years ago than it does now. The main issue then was to know whether the gravity-induced slow-down of the expansion was enough to stop it and produce a cosmic collapse or not; in this case, the expansion would continue forever. Much more powerful telescopes are operating now than decades ago, there are space telescopes producing breathtaking pictures, and new windows have been opened (infrared, radio, X-ray astronomy, gravitational waves). As a result, the optimism about our ability to understand the Universe according to the laws established in the Laboratory has evaporated gradually. Vera Rubin studied the rotation curves of the nearby Galaxies. She discovered that the stars in the outskirts

rotate a lot faster than they should if they were attracted by the visible matter only. There are also similar problems with Galaxy clusters and Fritz Zwicky in 1933 had already proposed the existence of Dark Matter . There is no agreement about the nature of this stuff, which is actively sought for. Stones are not candidates because they would alter the abundance of elements. WIMPS (weakly interacting massive particles) have long been predicted and sought, but never found. Black holes could do the job, perhaps.

8.10.4 Accelerated Expansion

Even more astonishing was the discovery that the expansion actually seems to accelerate. The accelerated expansion was discovered in 1998 and was quite unexpected. Two independent projects, the Supernova Cosmology Project and the High-Z Supernova Search Team, both used distant type Ia supernovae as standard candles, hoping to measure a deceleration due to gravity. In summary, very high redshift supernovae appeared to have a larger luminosity distance than expected from the Standard Model. This suggests an accelerating Universe for the following reason. The light has traveled a greater distance in an accelerating Universe than it would if the expansion rates were the same as in our neighborhood. Thus, for the same redshift, a supernova is more distant in an accelerating Universe, and hence, dimmer, than in other non-accelerating Universe models. This conclusion restored the cosmological constant Λ.

The cosmological constant does not change the symmetry of the model, thus the metric tensor remains diagonal. Thus, Λ can be fitted into the above Standard Model. One must increase the density by $\frac{\Lambda c^2}{8\pi G}$ and decrease the pressure by $\frac{\Lambda c^4}{8\pi G}$. So, Λ produces an extra energy, known as *dark energy*, associated with a negative pressure.[25]

8.10.4.1 Dark Energy

In 1968, YB Zel'dovich proposed an explanation of the cosmological constant Λ in terms of vacuum energy. The idea of an energy content of vacuum comes from Quantum Mechanics and will be developed in Sect. 16.2. Zel'dovich argued that in Special Relativity, the energy–momentum tensor of the vacuum should be proportional to $\eta_{\mu\nu}$ which is a tensor that keeps the same form under Lorentz transformations. Indeed the vacuum should appear the same in all inertial systems. By the principle of General Covariance this implies that in General Relativity $T_{\mu\nu}$ must be proportional to $g_{\mu\nu}$ and produce a Λ term in the Einstein equations. In this way, the idea of a dark energy

[25] However,the model with Λ has a competitor, known as Quintessence. This is thought to be a hypothetical scalar field; it should provide a dynamic dark energy in the sense that it generally has a density and equation of state that varies through time and space.

that becomes more and more dominant as the space grows could find an explanation. At the present time, Cosmologists suppose that the total mass–energy of the Universe contains 4.9% ordinary matter, 26.8% dark matter and 68.3% dark energy. It is embarrassing to admit that nobody is sure about the nature of dark energy and dark matter; but without them, the physical laws discovered in the laboratory and successful in the neighborhood of the solar system appear to fail.

8.10.4.2 Inflation

Even the isotropy of the Universe is somewhat paradoxical. This is the so-called horizon problem. How can we explain the fact that the cosmic microwave background follows Planck's law, and the temperature is the same (with minor fluctuations) in all directions? This implies that (1) there was equilibrium during the Big Bang, and (2) there was equilibrium even between parts of the Universe which had not yet exchanged radiation. Moreover, equilibrium implies maximum entropy, while, according to Thermodynamics, the entropy of the Universe is increasing. There is also a flatness problem: why is the Universe (within a few percent) flat? (Of course, if a small positive or negative curvature is discovered, the same people will invoke the curvature problem. Every aspect of the reality can be formulated as a problem, if we refuse to have initial conditions).

In order to avoid initial conditions, in 1980 Aleksej Starobinskij in URSS and Alan Guth in USA proposed the conjecture of an inflation field called inflaton that should have worked at the beginning of the Universe, for a duration of 10^{-30} s, starting 10^{-35} s after the big bang. (Note that currently one can measure 10^{-18} s.) So the Universe has been so small just long enough to produce thermal equilibrium. Once the task was done, this fantom field should be gone. So, we don't have the hassle of looking for it. Incredibly, this became popular among part of the scientific community. But it has unwanted features and has been questioned; the inflation field is an *ad hoc* assumption without any clear role in Physics. The explanation of phenomena should be in terms of established physical laws. The model removes the initial conditions but introduces an ad hoc unobservable field which needs initial conditions and acts in the first 10^{-35} s, but we do not know the laws of physics in such extreme conditions. It produces an expansion mechanism on top of the expansion already predicted in the standard model. Needless to say, it also produces unscientific unobservable mutiverses like ghost realities with random physical laws and has zero predictive power. This is not scientific, Penrose finds it horrible, and I agree. But at any rate, this topic is outside our scope.

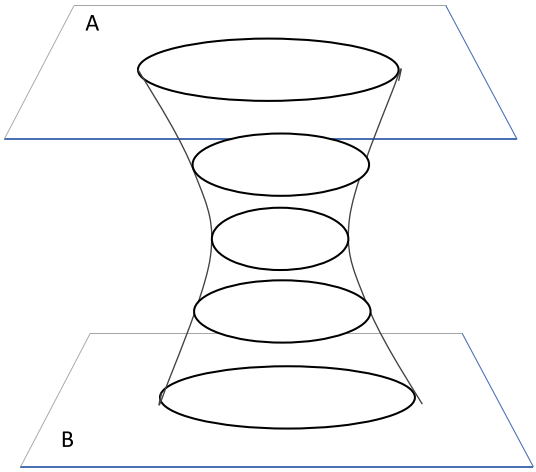


Fig. 8.6 A wormhole of the hypothetical kind (8.6) joining two regions of space–time

8.11 Conjectures About Wormholes and Time Machines

Instead of starting from the energy–momentum tensor and solving the Einstein equations to find the metric, one can start from the metric and seek the situation that could generate it. Wormholes may be exemplified by a metric of the form

$$d\tau^2 = dt^2 - dl^2 - (b_0^2 + l^2)(d\theta^2 + \sin^2(\theta)d\phi^2). \tag{8.97}$$

This represents, in obvious notation, two planar space–times A and B connected by a throat of radius b_0 through some extra dimension. An observer can go around in each of two flat space–times or go through the short throat and reach the other side (see Fig. 8.6).

Now, depending on the shape of our space–time in the augmented space, one can imagine that the two sides joined by the short throat are too very far away from each other, say in different galaxies, and so the wormhole is a short cut to places that are millions or billions of light years away. This situation is even more exciting if one considers that it also offers the possibility of joining two different times at the same place, that is, building time machines. The problem then is to look for the matter distribution that produces this exciting situation. The result[26] is that one needs negative energy or equivalently negative mass. Experimentally, one can produce a little negative energy density by the Casimir effect (see Chap. 16) but only locally, by shifting some zero-point oscillation of the electromagnetic field, but the idea of producing a wormhole in this way looks highly unphysical.

[26] Morris and Thorne (1988).

8.12 A Fleeting Glimpse Into the Einstein–Schrödinger Unified Field Theory, Kaluza's Five-Dimensional Conjecture and the Brans–Dicke Theory

In 1915, when General Relativity was published, nobody had any idea about the weak and strong interactions. The Einstein theory deals with gravity and electromagnetism, and does so successfully, by including the stress–energy tensor of the electromagnetic field of Sect. 6.4 as an additional contribution to the stress–energy tensor in Einstein's field equations of Sect. 8.4; as we know, light propagates in space–time following a null geodetic. The equation of motion of a charged particle reads

$$\ddot{x}^{\beta} + \Gamma^{\beta}_{\lambda\alpha}\dot{x}^{\lambda}\dot{x}^{\alpha} - \frac{q}{mc}g^{\beta\mu}F_{\mu\nu}\dot{x}^{\nu} = 0. \tag{8.98}$$

However, some theorists found aesthetic objections[27] to this scheme, motivated by the fact that the description of the gravity field was geometric and they would have preferred a unified description in which both fields had a geometric nature. These theorists (including Einstein) set out to find a united field theory. Einstein and Schrödinger thought that the symmetric g tensor of gravity should be combined with the antisymmetric electromagnetic tensor to produce a more general nonsymmetric metric tensor allowing to treat both fields on equal footing.

In 1921 Kaluza proposed to generalize Eq. (8.4), namely, $-ds^2 = g_{\mu\nu}dx^{\mu}dx^{\nu}$, by adding a fourth space coordinate x^4 and writing:

$$d\sigma^2 = \sum_{i,k=0}^{4} \gamma_{i,k}dx^i dx^k \tag{8.99}$$

where $k = 4$ is the "new" coordinate and all the $\gamma_{i,k}$ coefficients are taken to be independent of the newly assumed fourth space coordinate x^4. But that is too general, and Kaluza chose the appropriate five-dimensional metric. So, equation (8.4) is replaced by

$$ds^2 = g_{\mu\nu}dx^{\mu}dx^{\nu} + a(dx^4 + bA_{\mu}dx^{\mu})^2. \tag{8.100}$$

A long calculation of the geodetic in the five-dimensional space yields the correct equation of motion. One recovers the same Eq. (8.98) as General Relativity, with a suitable choice of the new parameters, but, so to say, *geometrically*.[28]

Now let us pause a moment to evaluate this approach physically. How reasonable is it to force electromagnetism and gravity in the same shape?

---

[27] Aesthetic criteria are arbitrary: they can give hints some time but Nature is not forced to follow them. For instance, the violation of Parity and Time reversal symmetries are unsightly things.

[28] Details may be found on the first Volume of the beautiful "Introduction to Modern Theoretical Physics" by Edward G. Harris.

While gravity is described by a nonlinear tensor theory, electromagnetism is a linear vector interaction (attractive as well as repulsive). This was well known already in 1921. Today we know that electromagnetism is a gauge interaction like weak and strong interactions and we speak about electroweak theory. Although these two forces appear very different at low energies, the quantum mechanical Weinberg–Salam theory shows that above the unification energy, on the order of 246 GeV, corresponding to a temperature of about 10^{15}○, there is a single electroweak force. In view of the existence of an electroweak force and the absence of an electrogravity force, the idea of generating electromagnetism by an extension of Relativity does not appear natural. Descriptions which are not natural but artificial are always possible. Mathematics allows everything, if one complicates the description enough; one can expand elephants in plane waves, but plane waves are not the best approach to understand elephants.

By the Kaluza modification of Relativity we gained no new Physics—actually our starting purpose was to get exactly the same physics with a more geometric approach, in terms of a geodetic. Kaluza succeded to do that. However, we have a terrible cost, which is the counterfactual assumption of a fourth space dimension. Klein suggested that the new dimension may be compactified and unobservable, but then we have an unobservable dimension, which is quite a bad assumption. A scientific theory should not make use of arcane interactions with undetectable entities, if we do not want to fall into witchcraft. Since the two interactions are so different, it seems much more reasonable that we treat them differently and do without the fifth dimension. This sterile theory has been considered outdated for a long time, because of the discovery of Quantum Mechanics and of two more interactions and because of the gigantic exciting progress achieved in such investigations.

But fifty years ago or so, string theorists resumed the idea of compactified dimensions, assuming a large (and adjustable) numbers of them. Despite the assumption of extra dimensions and super symmetry (another counterfactual assumption) there has been no scientific output from such conjectures. In my opinion, the total indifference toward the experimental data of a part of the community is very worrying.

On the contrary, the Brans–Dicke theory, published in 1961, can be considered as an extension of General Relativity in which the inverse $\frac{1}{G}$ of the gravitational constant is assumed to be a scalar field ϕ function of space and time. Having one more degree of freedom, the Brans–Dicke theory admits more solutions than General Relativity and could be a spare card should Einstein's theory need a fix. Since there are no known examples we may say that both theories agree with the experiment.

References

Bhadra, A., Ghose, S., Raychaudhuri, B.: (4 Aug 2022). arXiv:2207.09138v2

Dyson, F.J.: The world on a string. In: Greene, B. (ed.) Review of The Fabric of the Cosmos: Space, Time and the Texture of Reality, vol. 51, issue 8. New York Review Books (13 May 2004)

Everitt, C.W.F., et al.: Class. Quantum Grav. **32**, 224001 (2015)

Morris, M.S., Thorne, K.S.: Am. J. Phys. **56**, 395 (1988)
Rothman, T., Boughn, S.: (2 Dec 2006). arXiv:gr-qc/0601043v3
Ruggiero, M.L.: Eur. J. Phys. **24**, 563–573 (2003)
Visser, M.: (15 Jan 2008). arXiv:0706.0622v3 [gr-qc]

Chapter 9
The Wave Function

This theory has many counter-intuitive features, but perhaps the most amazing aspect is the absolute need for complex numbers, those that their inventor Girolamo Cardano (1501–1576) called "fictitious numbers".

Among the fundamental constants of Physics, Planck's[1] constant h is an action [h = energy × time = angular momentum]. It is a tiny action, $h \sim 6,62619610^{-34}$ J s. Most often, one uses the notation $\hbar = \frac{h}{2\pi} = 1.05410^{-34}$ J s. All the phenomena in which h is non-negligible are *quantum phenomena*, while sometimes phenomena in which the characteristic actions are $\gg h$ may have *classical* aspects, which allow for a simpler description; however, all the microscopic mechanisms are fundamentally quantum, and many phenomena such as magnetism, superfluidity and superconductivity are just macroscopic quantum phenomena; their quantum mechanical description is far from trivial. Historically, the breakdown of classical Physics (after many triumphs) was urged by the clear inability of the latter to explain a few quantum effects that were known by 1900. This breakdown was shocking, because well-understood laws of Physics that had been validated by experiment in many ways were seen to lead to wrong conclusions in a class of phenomena quite unexpectedly.

Without h, one cannot understand the stability of atoms.[2] From electromagnetism, we know that a charge e must irradiate a power

[1] Max Planck (Kiel 1858–Gottinga 1947) German physicist, initiator of Quantum Mechanics, Nobel for Physics in 1918.

[2] In 1900, micelles as small as 10^{-4} cm could be seen by microscopes, and atoms are just 4 orders of magnitude smaller. A jump of 4 orders of magnitudes implied completely new concepts. Today, some theorists aiming at the "theory of everything" boldly extrapolate the known laws of Physics to the Planck scale, which is some 15 orders of magnitudes away from what we can experiment. I am not surprised if there are no results.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_9

$$W = \frac{dE}{dt} = \frac{2}{3}\frac{e^2}{c^3}a^2, \tag{9.1}$$

where E is the energy of the atom, and a is the acceleration; this law explains very well the working of an antenna. Now let us try to build an atomic model where the electron makes an orbit of radius r around the fixed nucleus. Substituting Newton's law $a = \frac{e^2}{mr^2}$, one gets $W = \frac{dE}{dt} = \frac{2}{3r^4}r_0^3\frac{e^2}{c^3}$, where $r_0 = \frac{e^2}{mc^2} \sim 2.8\ 10^{-13}$ cm is the so-called classical radius of the electron. Now, the atom in a normal state does not radiate, yet the electron feels a force and is accelerated. Moreover, the electron should fall in the nucleus in a short time. The electrostatic energy of the atom is $E = -\frac{e^2}{2r}$, so

$$\frac{dE}{dt} = -\frac{r_0 mc^2}{2r^2}\frac{dr}{dt}.$$

Equating this to W, one finds that $r^2\dot{r} = -\frac{4}{3}cr_0^2$. We can integrate this with the initial condition $r(0) = a_0$ with the result that $r^3(t) = a_0^3 - 4r_0^2ct$. The radius becomes zero at $t = 1.6 * 10^{-11}$ s. This is in very sharp contrast with the experiment! Classically, one understands the chemical bond even less. Atoms, molecules, nuclei, solids and elementary particle interactions are all quantum effects. Later it became clear that the quantum theory was needed to understand macroscopic phenomena as well and for a variety of applications, including lasers and electronics. Classically, the cohesive energy of metals cannot be explained.

It is easy to see that magnetism would be impossible classically. This was the result of the Bohr–van Leeuwen Theorem. The Lorentz force correctly describes the motion of charges. However, for a classical electron in a magnetic field, the canonical partition function reads

$$Z = \int d^3x \int d^3p \exp\left[-\frac{(p_x - eHy/c)^2 + p_y^2 + p_z^2}{2mK_BT}\right];$$

this should be the starting point of description of magnetic effects in solids, but a trivial shift of the p_x variable yields

$$Z = V(2\pi mK_BT)^{3/2}.$$

The field has disappeared! Thus, classical physics predicts that magnetism does not exist, in sharp contrast with the experiment. Superconductivity and superfluidity are also macroscopic quantum effects. By contrast, the predictions of Quantum Theory are extremely accurate. In no other field has one gained a similar degree of knowledge; in some atomic phenomena, like the Lamb shift, 10 digits precision has been achieved.

This book is devoted to the non-relativistic version of the theory, although the relativistic one is well known. We shall be able to understand many important facts in a semi-quantitative fashion.

9.1 Corpuscles and Waves in Classical Physics

In classical physics, there are two types of basic objects, corpuscles and waves. The wave and particle concepts are well distinct. A corpuscle is a particle that obeys the canonical equations and follows a deterministic trajectory; if there is no force, it goes straight, with constant speed. The waves are described as solutions of a well-defined linear partial differential equation; they may be propagating in a elastic medium, or appear as continuous fields. In the presence of obstacles, they give rise to phenomena of interference and diffraction. Newton thought that light was made of corpuscles. This idea was natural in view of Geometrical Optics, based on light rays. This knowledge already enabled the building of powerful telescopes and microscopes that produced extraordinary progress. But at the half point of the nineteenth century, this type of geometrical optics was recognized as a limiting case of physical optics, that is, Maxwell's theory. It holds when all material objects are large compared to the wavelength. One phenomenon that the corpuscular model of Newton could not explain is diffraction. Actually, it was known since 1665, when an Italian scientist Francesco Maria Grimaldi, a Jesuit priest, discovered that *"Light propagates or spreads not only in a straight line, by refraction, and by reflection, but also by a somewhat different fourth way: by diffraction"*. Thomas Young in 1800 presented a paper to the Royal Society about the fact that light is a wave phenomenon and in 1801 and 1803 he presented the experimental evidence of interference and diffraction.

A more modern version of the experiment is as follows. A monochromatic plane wave of wave vector $\overrightarrow{k} = (k, 0, 0)$ parallel to the x axis impinges on a screen with a slit for $y \in (-a, a)$ and then exposes a photographic plate. To fix ideas, we take a wave polarized along the z axis with electric field

$$\mathbf{E} = (E_x, E_y, E_z) = \mathrm{Re}(0, 0, Ee^{ikx}). \tag{9.2}$$

The frequency and polarization of the field do not change. However, the beam undergoes a *Fraunhofer diffraction*. This is borne out by the Maxwell equations and is evident for ka of order 1. On the photographic plate, no sharp picture of the slit appears, but rather the intensity is a function $I(\theta)$ of the angle of deflection:

$$dI \propto \frac{sin^2(ka\theta)}{(ka\theta)^2} d\theta.$$

This function is reminiscent of a representation (3.7) of Dirac's δ; the narrower the slit, the broader the diffraction pattern. Beyond the screen, the wave vector of the diffracted field makes a small angle θ with the x axis; actually, it has got a y component $|k \sin(\theta)| \sim \mathrm{k}\theta$ (Fig. 9.1).

Classical Physics explains these facts very well. If the plane incident wave has constant amplitude E (independent of y), the part of the front cut off by the slit must depend on y as $\theta(a^2 - y^2)$.

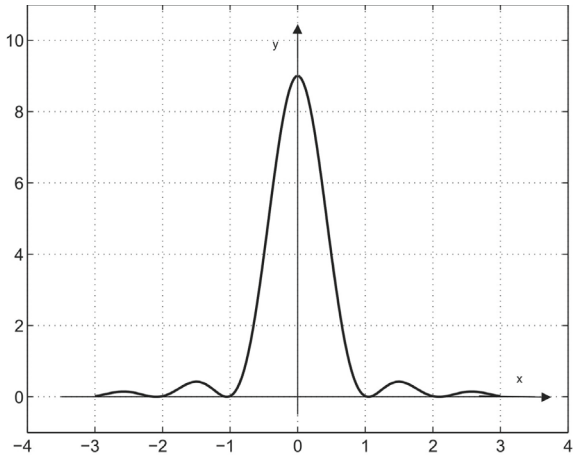


Fig. 9.1 Graph of $\sin^2(3x)/x^2$ near x = 0

Fourier's theorem gives the wave packet representation

$$\theta(a^2 - y^2) = \frac{1}{\pi}\int_{-\infty}^{\infty} dq e^{-iqy}\frac{\sin(qa)}{q}.$$

Letting $E \to E\theta(a^2 - y^2)$ in Eq. (9.2) one obtains

$$E_z = E\int dq e^{i[kx-qy]}\frac{\sin(qa)}{\pi q}.$$

A point of the plate at an angle $\theta \ll 1$ receives a component

$$E_z \propto e^{i[kx-k\theta y]}\frac{\sin\ (\mathrm{k}\theta a)}{\pi k\theta}$$

with a deflected wave vector, $\frac{q}{k} = \tan\theta \sim \theta$.

The intensity which is measured on the plate goes with the square of the field. Note carefully that

$$\boxed{E \text{ is an amplitude, the intensity is proportional to } |E|^2.}$$

Thus, it is clear that diffraction is a property of waves, and the amplitudes, not the intensities, are additive.

9.2 Dualism

If the intensity of the light is very weak, something happens that is not expected by Classical Physics. The pixels in the plate are exposed or not exposed in a yes or no fashion: there is no half measure. Higher and lower intensity of the radiation just changes the probability of each pixel of being exposed.

Max Planck introduced h in 1900 in order to fix the black body theory, affected by the *ultraviolet catastrophe*. We shall see the derivation of Planck's formula in the context of quantum statistical mechanics in Chap. 25 below. Then, in 1905 Albert Einstein understood that the light does not only travel as a bunch of quanta, but also is emitted and absorbed as quanta. This implies that it consists of quantized photons carrying each a quantized energy $E = h\nu = \hbar\omega$. This interpretation was forced by the *photoelectric effect*: experiments showed that the radiation picks electrons from a surface if its frequency is above a threshold, which depends on the material. The intensity of the electron current above the threshold is proportional to the intensity of the radiation. The relation between frequency and wave vector is $\omega = ck$, and the relation between energy and momentum is $E = cp$ for an electromagnetic wave, as a consequence of the well-known classical Poynting theorem. As a result of energy quantization, the photon momentum is also quantized and the momentum carried by a photon having wave vector k is $p = \hbar k$. We must accept the fact that a photon propagates as a wave, yet it is always absorbed as a corpuscle.

Then, in 1924, the French duke Louis-Victor De Broglie[3] suggested that in analogy with the dualism between photons and waves, particles like an electron should also correspond to a wave function ψ. Most remarkably, he also argued that ψ must be an unobservable, complex function

$$\psi_{\vec{p},E}(\vec{x},t) \approx \exp\left[\frac{i(\vec{p}\cdot\vec{x} - E(\vec{p})t)}{\hbar}\right], \tag{9.3}$$

with $\vec{p} = \hbar\vec{k}$ as the electron momentum. The square modulus is proportional to the intensity of the electron beam and to the probability of detecting the electron. The introduction of a complex quantity that is not observable but contains the physical information was a wonderful, bold flash of insight. In 1927, Clinton Davisson and L.H. Germer at Bell Labs USA confirmed De Broglie's hypothesis. (Clinton Davisson received the Nobel Prize in 1937). They fired electrons at a crystalline Ni target, and the resulting diffraction pattern was found to match the values predicted by the De Broglie formula. Since the late '50s, this is a technique known as Low Energy Electron Diffraction, used for the study of surfaces and their excitations. Here I use a more direct experiment to show the effect. This is feasible with the current technology but was impossible to do at the time (Fig. 9.2).

[3] De Broglie (1892–1987) won the Nobel Prize in 1929 for the very remarkable prediction about the existence and the form of a complex wave function for a free particle in a thesis work in 1924.

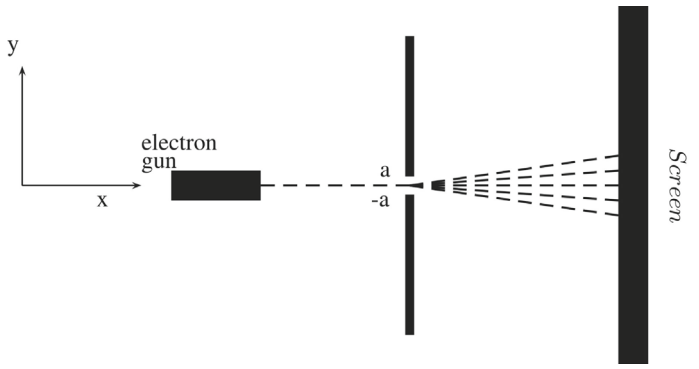


Fig. 9.2 Diffraction of an electron beam through a slit: the beam from gun C arrives at P in the detector plane (photographic plate, Geiger counter, etc.). The z axis is orthogonal to the plane of the figure

A beam of electrons of momentum **p** parallel to the x axis impinges on a screen and goes through a slit for $y \in (-a, a)$; the electrons meet a photographic plate (or better yet, a modern detector). The beam undergoes a Fraunhofer-like diffraction. The intensity at a small angle θ from the x direction turns out to go like

$$dI = \frac{\sin^2(ka\theta)}{(ka\theta)^2} d\theta.$$

This is not expected classically, but agrees with the De Broglie hypothesis (and indeed, no diffraction is seen if $ka \gg 1$). Again, if the incident wave is $\psi \propto e^{i\vec{k}\cdot\vec{r}}$, with initial wave vector $\vec{k} = (k, 0, 0)$, the transmitted wave is

$$\psi \propto \int dq e^{i[kx-qy]} \frac{\sin(qa)}{\pi q}.$$

For small angles, $|q| \approx k\theta$; evidently, the experiment confirms that

$$\boxed{\psi \text{ is an amplitude, and the intensity goes like } |\psi|^2.}$$

The wave function is the amplitude of something immaterial, a probability amplitude. In ordinary life, we can speak about evaluating the probability of an event when we have an incomplete information. (In the absence of any information one cannot estimate anything, and with complete information the outcomes are known). In Physical problems, we start with some informations, since we know how the experiment is set up. Quantum mechanics then allows to estimate the probabilities of different outcomes. One can never disentangle the properties of the system from what we know about it having set up the experiment in a particular way.

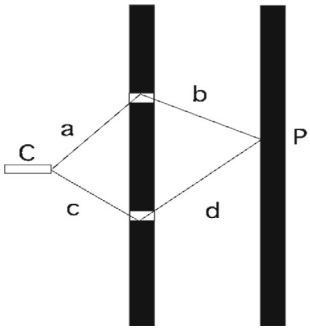


Fig. 9.3 The double slit experiment: the electron beam from gun C arrives to P on the photographic plate through two paths. The lengths are $l_1 = a + b$ and $l_2 = c + d$. An interference path is seen on the plate

In a one-dimensional problem, we may be considering a detection of a particle at some location **x**; the detection experiment can be done many times using a beam of particles; otherwise, we may do a series of experiments involving just one particle, like an electron in an H atom, but repeating the experiment on many samples. In both instances, it makes sense if we ask for the probability that the detection occurs in a volume V. This is given by

$$P_V = \int_V d^3x |\Psi(\mathbf{x}, t)|^2,$$

where I have included a possible dependence on the time t because the system could evolve. In other words,

$$\rho(\mathbf{x}, t) = |\psi(\mathbf{x}, t)|^2 \tag{9.4}$$

is a probability density that must be normalized to 1 by

$$\int d\mathbf{x}\rho(x, t) \equiv 1,$$

integrating over all space, since the particle must be somewhere.

The quantum theory does not propose a *compromise* between the conflicting views of waves and particles. It does not attempt any pun. It makes extremely precise predictions based on a rigorous mathematical apparatus which has the precedence over any argument. Wave-like diffraction and particle-like detection take place in the same experiment (Fig. 9.3).

9.3 Which Way Did the Electron Pass Through?

The double slit experiment with electrons (Fig. 9.2) is an even more striking display of the wave-like properties of matter particles. The intensity at a point P on the plate is proportional to

$$1 + \cos[k(l_1 - l_2)],$$

where l_1, l_2 are the path lengths. When one of the slits is closed, the interference pattern between the two waves disappears, but one observes the diffraction pattern of the other slit. This is like the double slit experiment familiar in classical physical optics, but interference and diffraction cannot be understood in terms of a beam of corpuscles. One could suspect that the electron–electron interaction has something to do with this; but, the intensity of the electron beam can be made so low that in each moment, it is very unlikely that more than one electron is present. Then, the experiment must be continued for a long time, but when enough electrons reach the plate, the same results as before are obtained. The pattern persists and is a single-electron property. Thus, the electron wave must interfere with itself! Moreover one can never tell which slit a given electron went through; all attempts to modify the experiment and determine the electron trajectory destroy the interference. Currently,[4] double slit experiments are also done with atoms. A mathematical description of the interference experiments is deferred to Sect. 15.2

9.3.1 The Bohm–Aharonov Effect

A flux tube is a solenoid that produces a magnetic field inside a cylinder where the electrons cannot enter, but the field is screened, for instance, by coating the solenoid with a superconducting film.[5] Therefore, there is a field inside but not outside. In 1959, D. Bohm[6] and Aharonov proposed a startling modification of the double slit experiment, using a flux tube, as shown in Fig. 9.4.

Therefore, the electrons move in a field-free region where there can be no Lorentz force; no effect of the field confined in the tube on the interference pattern can be expected classically. Nevertheless, there is a magnetic flux through a surface bounded by the path abdc. This induces a phase difference between the wave function along the path ab and the wave function along cd, which shifts the interference path on the screen (for the theory, see Sect. 15.2).

[4] See for instance A. Miffre et al., $arXiv/cond-mat$, 0506106.

[5] A superconductor is a perfect diamagnet and expels the magnetic field totally. We shall meet them in Sect. 26.5.

[6] David Bohm (1917–1992) was an American–Brazilian–British theoretician who, among other things, contributed to the Manhattan Project, worked on the EPR paradox and reformulated Quantum Theory.

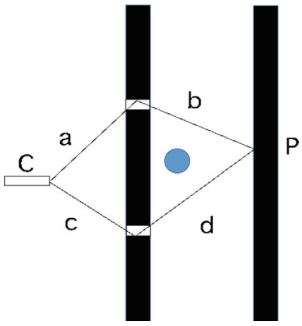


Fig. 9.4 The double slit experiment modified by a flux tube (round dot). The interference at P depends on the flux collected within the path abdc, even if there is no magnetic field outside the flux tube

9.3.2 Experiments by Deutsch on Photons

New enlightening experiments using photons have been performed by D. Deutsch et al.[7] The light from source S is split by the half silvered mirror SST (Fig. 9.5 left). Classically, detectors A and B should receive half of the intensity, as predicted by the Maxwell equations, and this happens if the intensity is such that the number of photons in the system at any given time is large. However one can decrease the intensity to one quantum per second, then no more than one photon is around. Then, the photon ends up either in A or in B, but with equal probability, so the classical prediction is verified only on average when many experiments have been conducted. But what happens each time?

It is natural to suppose that once in the half silvered mirror SST, the photon makes a random choice. Is that true? One can find an answer through the second experiment in Fig. 9.5 right, done with one photon at a time. The geometry is such that there is constructive interference in A and destructive interference in B. As a matter of fact, the lonely photon always goes toward detector A. We must conclude that (1) there is no random choice, and (2) the particle somehow explores both paths and interferes with itself! If one of the paths is slightly modified, the detector in B starts counting photons. This is really striking. So to say, the incoming photon is aware of the boundary conditions. At any rate, the quantum behavior supersedes the classical description at low enough intensity.

The classical field at high intensity is correctly predicted by the Maxwell equations. The next question is: what is the photon wave function? The wave function is not an observable quantity and as such, as we shall discover below, it is always somewhat arbitrary. The general reply is the four-potential, but here it is enough to take the vector potential $\overrightarrow{A}$ as the wave function since only the transverse field is involved. In the Lorentz gauge, the vector potential $\overrightarrow{A}$ obeys the wave equation; from $\overrightarrow{A}$, one can derive the electric field $\overrightarrow{E} = \frac{\partial \overrightarrow{A}}{\partial t}$, and the light intensity is proportional to $\overrightarrow{E}^2$. There is a partial wave in both paths, and all the intensity goes to detector A if there is constructive interference in A and destructive interference in detector B.

[7] See for instance David Deutsch, Arthur Ekert and Rossella Lupacchini, math.HO/9911150.

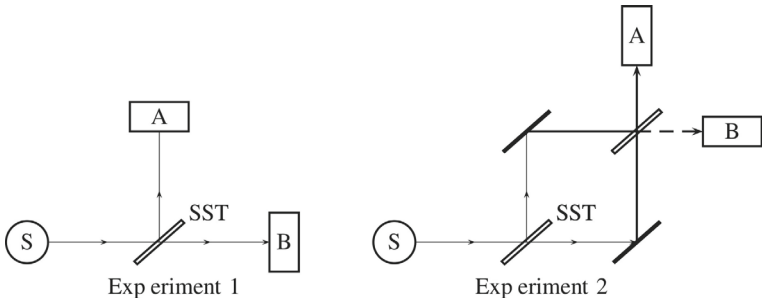


Fig. 9.5 In experiment 1, by a semi-transparent mirror, the photon is sent to A or B with equal frequency. In experiment 2, by a double semi-transparent mirror and two ordinary mirrors, the photon is sent to A

Even at low intensity, the wave equation is useful and the square of $\overrightarrow{A}$ is proportional to the probability of detecting the photon. Maxwell's equations survive unharmed in Quantum Mechanics, but must be interpreted in a new way. This interpretation is also consistent with the first experiment of Fig. 9.5. A more general implication of this example is that each species of particle has its own wave function.

But a further question arises naturally: how can we predict in the first experiment where the next photon will go? Here, the answer is: in no way! A similar experiment has been performed[8] with neutrons.

9.4 Quantum Eraser Experiments and Delayed Choice Experiments

In double slit experiments, the interference pattern vanishes if, by the use of light or by other means, one can in principle detect which slit the particle passes, but whenever this information is erased the interference pattern returns. There are other ways to study the interplay of particle-like and wave-like properties. Let us go back to Sect. 9.3.2 and to Fig. 9.5. In Experiment 1, a John Wheeler noticed some interesting implications of this experiment. In the latter case, the wave properties are evident, which implies that the photon goes both ways. In some sense, the photon behaves as a particle or as a wave depending on the presence or absence of the second beam splitter. One can say that without the beam splitter we can tell the path of the photon, but by introducing the second beam splitter, the path information has been *erased*. The cancelation produces interference phenomena at detection screens A and B positioned just beyond each exit port, where the count rate depends on the optical paths. However, it takes time to go from the source S to A or B. What happens if the

[8] Iannuzzi (2006).

second beam splitter is inserted into the left panel (or removed from the right panel) while the photon is in flight?

Experiment says that when the second beam splitter is present, the photon behaves as a wave even if it is inserted during the flight; if it is removed during the flight the photon behaves as a particle even if it must initially have started as a wave. In a sense, it is possible to change the *decision* of the photon in a retroactive way. This is actually an upsetting conclusion.

A cosmic version of the experiment is even more impressing. Sometimes, multiple images of a quasar are produced in the field of a Galaxy by gravitational lensing (Sect. 8.7). One can orient a telescope in such a way that only one of the images is visible; in this case, only the photons that according to one interpretation *decided* to travel as particles can contribute (left part of Fig. 9.6). However one can also produce an interference pattern out of two images (right part of Fig. 9.6) and in this case the photons must have *decided* to travel as a wave. Is it possible that the experimental arrangement on Earth reverses a *decision* made millions of years ago?

The situation is clarified by an enlightening paper published online by David Ellerman, entitled: "A Common Fallacy in Quantum Mechanics: Why Delayed Choice Experiments do NOT Imply Retrocausality". Essentially, the argument is that the photons do not have to *decide* their mode of propagation (particle-like or wave-like).

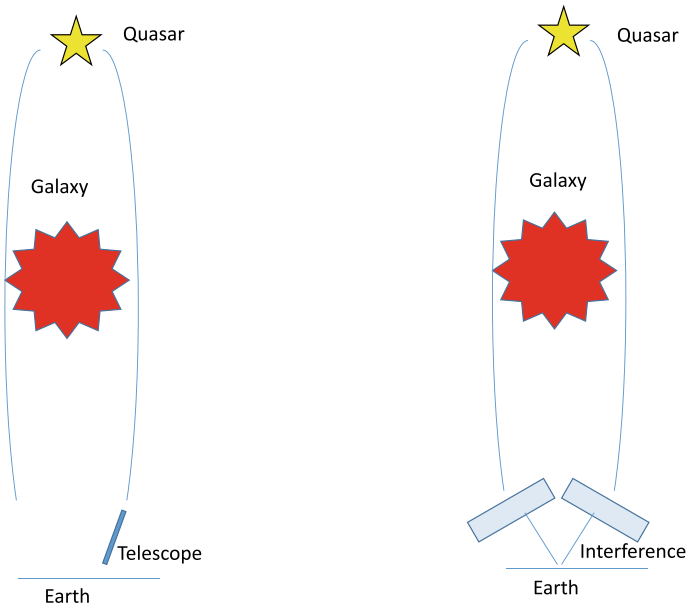


Fig. 9.6 Left: only one of the images of the quasar enters the telescope. Right: collecting both left and right images of the quasar and superposing them, we erase the path information and produce an interference pattern

Rather, let us go back to Fig. 9.5: the wave functions are a quantum superposition of propagation along the vertical path, propagation along the horizontal path and propagation along both paths; it is the choice of the method of detection that makes the wave function collapse in a particular way. In this way, the upsetting perspective of a violation of causality and/or retrocausality is removed.

9.4.1 Plane Wave, Superposition Principle and Operators

The experimental evidence discussed in the above sections strongly validates the intuition by De Broglie that a free electron having momentum $\overrightarrow{p}$ and energy E is represented by a complex wave with wave vector $\overrightarrow{k}$. The De Broglie wave of the form:

$$\psi_{\overrightarrow{p},E}(\overrightarrow{x},t) = \frac{1}{(2\pi)^{3/2}} \exp\left[\frac{i(\overrightarrow{p}\cdot\overrightarrow{x} - E(\overrightarrow{p})t)}{\hbar}\right], \tag{9.5}$$

with $\overrightarrow{p} = \hbar\overrightarrow{k}$, $E = \frac{\overrightarrow{p}^2}{2m}$; the normalization factor $\frac{1}{(2\pi)^{3/2}}$ ensures the orthogonality of De Broglie waves with different momenta, that is,

$$\int \psi^*_{\overrightarrow{p'},E}(\overrightarrow{x},t)\psi_{\overrightarrow{p},E}(\overrightarrow{x},t)d^3x = \delta^3(p-p'). \tag{9.6}$$

Actually, $\psi_{\overrightarrow{p},E}$ is interpreted as an amplitude, and since it is complex we cannot give it any direct physical meaning, while the square modulus represents an intensity of the wave, namely, $|\psi(\mathbf{x},\mathbf{t})|^2$ is the probability of finding the particle in $\mathbf{x}$ at time t. In the case of the free particle, $|\psi_{\overrightarrow{p},E}(\overrightarrow{x},t)|^2$ does not depend on the position; this means that we have no information concerning the position of the particle.

However, in the presence of fields, the wave function must be labeled by different quantum numbers in place of the momentum components and must be obtained by solving the Schrödinger equation, that we shall meet in the next chapter. The mean position of the packet must be computed as a statistical mean using the modified wave $\psi(\overrightarrow{x},t)$:

$$<\overrightarrow{x}>=\int d\overrightarrow{x}\,\overrightarrow{x}\,|\psi(\overrightarrow{x},t)|^2 = \int d\overrightarrow{x}\,\psi(\overrightarrow{x},t)^*\overrightarrow{x}\,\psi(\overrightarrow{x},t). \tag{9.7}$$

In quantum theory, the probability of detecting a particle in $\overrightarrow{x},t$ (that is, at $\overrightarrow{x}$ at time t) must be proportional to $|\psi_{\overrightarrow{p},E}(\overrightarrow{x},t)|^2$. Using (9.5), this is constant, so actually we have no idea of where the particles are. Indeed, the De Broglie wave function is a very special case, appropriate when we know momentum and kinetic energy of the particle, but we have no information at all about its position; in fact, the integral (9.7) diverges in this case.

If the reader finds that this information looks very incomplete, compared to what one would deem a full classical description, I cannot deny that this is true. In Quantum

Mechanics, like in electromagnetism, the superposition principle is valid, and a linear combination of wave functions is a possible wave function.

A De Broglie wave function $\psi_{\overrightarrow{p},E}$ represents particles prepared in a state with well-defined momentum,[9] so it contains the information that the momentum is $\mathbf{p}$. How can we get the momentum $\mathbf{p}$ out of it *by means of a linear operation*? A logarithm cannot work, because of the linearity constraint. However, we can introduce the *momentum operator*

$$\overrightarrow{p}_{op} = (-i\hbar)\overrightarrow{\nabla}, \tag{9.8}$$

which gives us

$$\overrightarrow{p}_{op}\psi_{\overrightarrow{p},E}(x) = \overrightarrow{p}\,\psi_{\overrightarrow{p},E}(x). \tag{9.9}$$

Mathematically, this is a well-known equation, an eigenvalue equation: the proper statement is that $\psi_{\overrightarrow{p},E}$ is an eigenfunction of the momentum operator and $\overrightarrow{p}$ is the corresponding eigenvalue. This complex function represents a particle moving in the direction of $\overrightarrow{p}$. We can get a real wave function by summing $\psi_{\overrightarrow{p},E}(\overrightarrow{x},t)$ with $\psi_{-\overrightarrow{p},E}(\overrightarrow{x},t)$. The opposite currents sum to zero. We shall see in the chapter on Quantum Transport that a real wave function cannot carry a current.

It is time to better specify what information can be extracted from ψ. Sometimes, broadly speaking, we say that ψ describes the particle. Actually, the average over the wave function tells us about the result of a series of many measurement on the sample containing many identically prepared particles, rather than the properties of a single particle. One can choose experimental conditions such that only one particle is present. Then, ψ cannot predict the position of the particle. But if the single-particle experiment is repeated many times, then ψ can be used to make statistical predictions. In a plane wave state, as we have seen, the position of the particles is quite undetermined. A general state is a superposition called a wave packet[10]:

$$\psi(\overrightarrow{x},t) = \frac{1}{(2\pi)^{3/2}}\int d\overrightarrow{p}\; g(\overrightarrow{p})\exp\left[\frac{i(\overrightarrow{p}\cdot\overrightarrow{x} - E(\overrightarrow{p})t)}{\hbar}\right]. \tag{9.10}$$

This is a superposition of De Broglie waves.

Here, $g(\overrightarrow{p})$ is a complex probability amplitude (its square modulus is the probability) that a momentum measurement yields $\overrightarrow{p}$. In effects, $g(\overrightarrow{p})$ is the *wave function in the momentum representation.*

If the absolute value of $g(\overrightarrow{p})$ is strongly peaked around some value $\overrightarrow{p}_0$, there is a large uncertainty about its location. At the limit, when the momentum is sharply defined, we are back to the De Broglie wave.

The information contained in $g(\mathbf{p},t)$ is equivalent to the knowledge of $\psi(\mathbf{x},t)$ and we can treat $g(\mathbf{p},t)$ as a wave function in the *momentum representation.*

---

[9] Ideally, a good electron gun can prepare a sample of electrons having rather well-defined energy and momentum.

[10] The wave function must always have a Fourier transform in the space of ordinary functions or in the space of distributions.

Problem 9.1 Prove that

$$< \overrightarrow{p} >= \int d^3\overrightarrow{p}\,\overrightarrow{p}\,|g(\overrightarrow{p})|^2; \tag{9.11}$$

yields the same result as

$$\int d^3\overrightarrow{x}\,\psi(\overrightarrow{x},t)^*(-i\hbar)\overrightarrow{\nabla}\psi(\overrightarrow{x},t) =< \overrightarrow{p} > \tag{9.12}$$

and so represents the mean value of the momentum.

Solution 9.1 Differentiating equation (9.10), one obtains

$$-i\hbar\nabla\psi(\overrightarrow{x},t) = \frac{1}{(2\pi)^{3/2}}\int d^3\overrightarrow{p}\; g(\overrightarrow{p})\frac{\overrightarrow{p}}{\hbar}\exp\left[\frac{i(\overrightarrow{p}\cdot\overrightarrow{x} - E(\overrightarrow{p})t)}{\hbar}\right],$$

and from (9.10) we find that

$$\psi(\overrightarrow{x},t)^* = \frac{1}{(2\pi)^{3/2}}\int d^3\overrightarrow{p'}\; g^*(\overrightarrow{p'})\exp\left[\frac{-i(\overrightarrow{p'}\cdot\overrightarrow{x} - E(\overrightarrow{p'})t)}{\hbar}\right]. \tag{9.13}$$

Setting up the integral (9.12) and integrating over $d\overrightarrow{x}$, using

$$\int d^3\overrightarrow{x}\,e^{i[\frac{(\overrightarrow{p}'-\overrightarrow{p})}{\hbar}]\cdot\overrightarrow{x}} = (2\pi)^3\hbar\delta(\overrightarrow{p} - \overrightarrow{p}'),$$

one finds

$$\int d\overrightarrow{x}\,\psi(\overrightarrow{x},t)^*(-i\hbar)\overrightarrow{\nabla}\psi(\overrightarrow{x},t) = \int d\overrightarrow{p}\,d\overrightarrow{p'}\,\delta(\overrightarrow{p} - \overrightarrow{p'})g^*(\overrightarrow{p'})g(\overrightarrow{p})\overrightarrow{p},$$

which is equivalent to (9.11). We see that we needed the factor $\frac{1}{(2\pi)^{3/2}}$ introduced in (9.10).

This example shows how the average momentum in the coordinate representation $\psi(x)$ is obtained by the **momentum operator** introduced in (9.8),

$$\overrightarrow{p}_{op} = (-i\hbar)\overrightarrow{\nabla}.$$

The classical variables become quantum operators; the way the operators are written depends on the representation. We have already seen that when acting on a $\psi(x)$, the correspondence is $\overrightarrow{x} \to \overrightarrow{x}_{op} = x$, $\overrightarrow{p} \to \overrightarrow{p}_{op}$. In general, if A is a physical quantity (that is something that can be measured) we call it an *observable*. There are quantum observables (like the spin of electrons, the lepton number and the

flavor of quarks) that do not correspond to classical variables. For each observable, there is an operator $\hat{A}$ such that

$$< A >= \int d\overrightarrow{x}\, \psi(\overrightarrow{x}, t)^*\, \hat{A}\, \psi(\overrightarrow{x}, t)$$

is the average value of many measurements. The De Broglie wave (9, 5) is an example of a general rule: the eigenfunction of $\hat{A}$ with eigenvalue a is denoted by ψ_a; if the physical magnitude corresponding to $\hat{A}$ is measured, the result is granted to be the eigenvalue a. It also gives a hint as how to find the operator of kinetic energy $\hat{T} = \frac{\mathbf{p}^2}{2m}$. Similarly, the angular momentum operator $\hat{L} = \mathbf{r} \wedge \mathbf{p}$, which depends on coordinate and momenta, is obtained letting $\mathbf{p}$ become $(-i\hbar)\overrightarrow{\nabla}$. However, in general caution is necessary in transforming classical observables into operators, since quantum operators may fail to commute, as we shall see next.

Commutator

Using a test function $\phi(x)$, one can observe that the commutator $[p, x]_-\phi(x) \equiv (px - xp)\phi(x) = -i\hbar\phi(x)$ does not vanish. The operators $\hat{x}$, $\hat{p}$ do not commute:

$$\boxed{[p, x]_- = px - xp = -i\hbar.}$$

This *fundamental commutation rule* replaces the classical fundamental Poisson bracket. We shall see that this analogy is profound.

Example 9.1

$$(x - ip)(x + ip) = x^2 + p^2 + \hbar.$$

9.5 The Collapse of the Wave Function and the Copenhagen Interpretation

Quantum Mechanics is a physical theory, and its purpose is the description of Nature and the prediction of the outcome of experiments. The theory is so successful that everyone believes it must be true. It allows us to predict the results of all experiments done so far in particle, nuclear, atomic, molecular and condensed matter Physics so well that no failure of Quantum Mechanics has ever been reported.

The fundamental laws of Nature are so different from the picture of reality that we form in the everyday life that we feel confused. The great Richard Feynman declared that *nobody understands Quantum Mechanics*; a bold way to say that the classical intuition is misleading in the quantum world. It is not just the random character of the results of some measurements; that is a fact, which requires a statistical treatment. The trouble is that everyone would like to infer from the experimental results a cartoon of what really happens at the microscopic level, and this requires a totally new way of thinking about particles and interactions. In Physics, the accuracy of

measurements and the quantitative predictions of the theories are of fundamental importance, but qualitative arguments are equally essential. This is what we mean by interpretation.

The use of probability theory had appeared in Theoretical Physics when Gibbs invented Statistical Mechanics. Here, the meaning is quite different and has nothing to do with ensemble averages, while the Temperature is not involved.[11] In Classical Physics we need statistics when we have incomplete knowledge of a system. Knowledge of the internal state of a slot machine enables us to predict the right moment for inserting the coin; since the internal state is secret, one can only speak about probability. Is it possible to go beyond the bizarre rules of the game and fill the gap between what appears from the measurements and the underlying reality?

In the *Copenhagen interpretation,* proposed by Niels Bohr and Werner Heisenberg in the 20s, the reply to the last question is a radical no. The wave function ψ contains a statistical description, which, however, is a full description of reality. The particle does not possess such properties as a classical trajectory. The only values that can be given to observables are those of operators for which ψ happens to be an eigenfunction. The microworld is so unusual that a particle can have many non-classical properties, including spin, isospin, parity, strangeness and color charge; on the other hand, particles lack familiar properties like a sharply defined trajectory. In the case of the slot machine, the missing information is hidden, but the precise position of the electron prior to measurement is fuzzy.

Suppose somebody wants to measure the electron–proton distance in the normal state of the H atom, by sending fast charged particles and measuring the change of their momentum, which depends on the field they come across. The measurement on a single H atom can give any value. Preparing many atoms in the same conditions and repeating the measurement many times, one gets a well-defined distribution of electron–proton separations. The result can be interpreted as a $|\psi(\overrightarrow{r})|^2$ and allows us to calculate the statistical average $\langle r \rangle = \int |\psi(\overrightarrow{r})|^2 d^3r$. This is not the same as repeating the measurement many times on the same atom, because the measurement itself perturbs the system, while classically, this kind of complication is assumed to be negligible. More precisely, the measurement of an observable $\hat{O}$ yielding the result o leaves the system in an eigenstate ψ_o of $\hat{O}$ with exactly that eigenvalue.[12] In other words, the wave function has collapsed to ψ_o. The *collapse of the wave function* is one of the most striking predictions of Quantum Mechanics and leads, among others, to paradoxes to be discussed in Chap. 27.

[11] Indeed there is a link between Quantum and Statistical Mechanics, which are unified by now, e.g., in the Keldysh theory. This exploits the analogy between the Boltzmann factor and the exponential $\exp \frac{-iEt}{\hbar}$ which appears, e.g., in the De Broglie wave and in stationary states (see below). However, conceptually these remain quite different exponentials.

[12] If the measurement is repeated afterward the system is found in a state $\psi_o(t)$ that is evolved from ψ_o, and if $\hat{O}$ commutes with the Hamiltonian, the eigenvalue is still o. This statement may be somewhat obscure to the reader but will be clear after studying the evolution in the next chapter.

There are also quantities that are sharply defined under some circumstances. If a sample of H atoms in the ground state is used to measure the binding energy, the result is about 13.59 eV and is always the same within the experimental error.

In general, however, one must make an average over an appropriate distribution. In the simplest cases, we have a probability density like Eq. (9.4). Let us make some examples in one dimension. One obtains the *mean value* of any function $f(x)$ defined by

$$\langle f(x)\rangle = \int_{-\infty}^{\infty} dx\rho(x,t)f(x). \tag{9.14}$$

This is commonly referred to as the *expectation value*, even if everybody understands that this is a misnomer. For instance, a probability distribution symmetric around x_0 could vanish right there, so x_0 would be unattainable but still the expectation value.

The average position at time t is

$$\langle x(t)\rangle = \int_{-\infty}^{\infty} dx\rho(x,t)x, \tag{9.15}$$

which implies

$$\langle\{x - \langle x\rangle\}\rangle = \int_{-\infty}^{\infty} dx\rho(x,t)\{x - \langle x\rangle\} = 0. \tag{9.16}$$

As in every statistical treatment, the width of the distribution is given by the standard deviation σ defined by

$$\sigma^2 = \langle\{x - \langle x\rangle\}^2\rangle = \int_{-\infty}^{\infty} dx\rho(x,t)\{x - \langle\, \mathrm{x}\rangle\}^2\,. \tag{9.17}$$

One finds that

$$\sigma^2 = \langle x^2\rangle - \langle x\rangle^2, \tag{9.18}$$

where

$$\langle x^2\rangle = \int_{-\infty}^{\infty} dx\rho(x,t)x^2. \tag{9.19}$$

The wave function is complex and can be taken to be real only in special cases. This is far from obvious since the measurements always yield real results. However, we have seen that the De Broglie wave is complex and is the eigenfunction of momentum, which is itself a complex operator. We shall see that the phase is crucial for angular momentum eigenstates, for the existence of a current, for enabling a change of gauge and a change of reference in a Galileo transformation; later on, we shall discover more subtle ways the phase is necessary (topological phase). On the other hand, the wave function, like the electromagnetic potentials, is not observable.

In simple systems like a particle in a quantum well, one measurement of energy can tell us which one is the state. In general, the task of labeling a system with a quantum state can be demanding. Complex systems with several degrees of freedom may require a series of measurements. How many quantities and which quantities are needed? This complicated issue is studied by *Quantum Tomography.*

Reference

Iannuzzi, M., et al.: Phys. Rev, Lett (2006)

Chapter 10
The Schrödinger Equation and the Evolution Operator

It is a strange theory that predicts particles that interfere with themselves and propagate through forbidden regions. But that is just what happens in reality.

10.1 The Schrödinger Equation

The years 1925–1926 saw the great turn. While Pauli solved the Hydrogen atom model using the Runge–Lenz vector and Heisenberg was developing his matrix method, the Viennese Erwin Schrödinger developed the fundamental equation. He was a full professor and an expert in boundary value problems; after the De Broglie thesis proposing a plane wave function for the electron, it was natural for him to look for a differential equation for the wave function. His fundamental work "Quantisierung als ein eigenwert Problem" appeared in Annalen der Physik in 1926 and was immediately successful. To this end, he had to devise the operators that represent variables of classical dynamics; the familiar ones are built by analogy with classical analogues,[1] but the momentum operator is given by Eq. (9.8),

$$\overrightarrow{p} = (-i\hbar)\overrightarrow{\nabla}.$$

[1] Since the classical description, as we know, can be changed by canonical transformation, this statement implies that in Quantum Mechanics, we must enjoy the same freedom. We shall see later how this arises.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_10

In the trivial case of free particles, the De Broglie wave (9.5) obeys[2]

$$\hat{\vec{p}}\,\psi_{\vec{p},E} = \vec{p}\,\psi_{\vec{p},E}, \tag{10.1}$$

moreover, since $\hat{H} = -\frac{\hbar^2}{2m}\nabla^2$, it follows that

$$\hat{H}\psi_{\vec{p},E} = E\psi_{\vec{p},E}. \tag{10.2}$$

Note that $\psi_{-\vec{p},E}$ has the same energy as $\psi_{\vec{p},E}$, so any linear combination of the two still has energy E.

Inspired by the De Broglie wave, where the energy is at the exponent, Schrödinger decided that the energy operator $\hat{E}$ must be

$$\hat{E} = i\hbar\frac{\partial}{\partial t}. \tag{10.3}$$

Indeed, the eigenvalue equation

$$\hat{E}\psi_{\vec{p},E} = E\psi_{\vec{p},E}, \tag{10.4}$$

where E is the energy, holds. In time-independent problems, the energy must be conserved and classically $H = E$.

Therefore, Schrödinger proposed his equation;

$$\boxed{\hat{H}\psi = \hat{E}\psi.} \tag{10.5}$$

Generally of course, there are complications due to the fact that some operators fail to commute. For the simple case of one particle in a potential,

$$\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + V(\hat{x}). \tag{10.6}$$

The position is an operator, now. Indeed, V is also an operator, namely, the operator that multiplies by V, but since it makes just a multiplication the hat is often omitted.

By the same criterion, one can write down the Hamiltonian operator for systems of particles.

Actually, the Schrödinger equation (10.5) is one of the most important equations of history and describes the evolution of single-particle and (involving more principles, that we shall see later) many-particle systems too (see Chap. 19). It remains valid

[2] In order to stress that a symbol represents an operator that acts on a wave function, one may use a hat (like in the case of the Hamiltonian $\hat{H}$ in this case).

also for problems which are not stationary; actually it dictates the evolution of all states. If H is time-independent, we may seek wave function with sharp energy that are eigenstates of both H and $\hat{E}$.

For a particle in a (possibly time-dependent) potential $V(x,t)$ in one space dimension, one must solve this linear partial differential equation:

$$\left[\frac{\hat{p}^2}{2m} + V(\hat{x},t)\right]\psi(x,t) = i\hbar\frac{\partial}{\partial t}\psi(x,t), \tag{10.7}$$

where $\hat{p} = -i\hbar\frac{d}{dx}$; the extension to three dimensions is immediate, with $\hat{p} \to -i\bar{\nabla}$.

It is common to speak about waves and wave mechanics, but the Schrödinger equation is no wave equation; actually it becomes a diffusion equation after the transformation $t \to i\tau$.

It proves to be an excellent idea to introduce the *time evolution operator* $\hat{U}(t,t')$ such that

$$\psi(x,t) = \hat{U}(t,t')\psi(x,t'); \tag{10.8}$$

the evolution operator then obeys

$$\left[\frac{\hat{p}^2}{2m} + V(\hat{x},t)\right]U(t,t') = i\hbar\frac{\partial}{\partial t}U(t,t'). \tag{10.9}$$

Perhaps, this way to rephrase the evolution in terms of operators does not immediately appear to be a clever idea; it proves very useful in the development of the theory, since $\hat{U}$ has several interesting properties that have been exploited with much ingenuity. To start with, the formal solution of Eq. (10.9)

$$U(t,t') = 1 - \frac{i}{\hbar}\int_0^t H(t_1)U(t_1,t) \tag{10.10}$$

lends itself to an iterative expansion. The so-called Group property

$$U(t,t_a)U(t_a,t') = U(t,t') \tag{10.11}$$

which is obviously true for any t_a is the starting point of the Path Integral formalism, as I show below (Sect. 15.2).

For time-independent problems, one can write

$$U(t,t') = \exp\left[-i\frac{H(t-t')}{\hbar}\right]. \tag{10.12}$$

When $\psi(\hat{x}, t=0)$ is known, (10.7) determines its evolution in the future and in the past. Linearity implies the superposition principle. Equation (10.5) is second-order in the space derivatives and first-order in time, so it breaks the principle of Relativity. Already in the present form, it explains a lot of phenomena. The equation can be used

as such for electrons, provided that spin does not enter into play. Moreover, it was the starting point for the relativistic extension of Quantum Mechanics (Klein–Gordon, Dirac, Proca equations).

10.2 The Ehrenfest Theorem

The change from Classical Mechanics to Quantum Mechanics is far from obvious and the correct interpretation requires a careful thought; here we are primarily interested in the scientific aspect, leaving philosophy to philosophers. Due to the presence of the Hamiltonian in the Schrödinger equation, it is clear that the new theory is related to Classical Mechanics; however, Paul Ehrenfest made the relationship much more explicit by showing that for any operator $\hat{A}$,

$$\frac{d\hat{A}}{dt} = \langle \frac{\partial \hat{A}}{\partial t} + \frac{[\hat{A}, \hat{H}]}{i\hbar} \rangle. \tag{10.13}$$

This is the Ehrenfest Theorem and shows that the evolution of the mean values of any observable is the same as the classical evolution with the mean value of the force (which, however, is not the same as the classical force, since the force depends on position, and the position is not sharp). In other terms, we are getting a wave instead of the familiar trajectories, but at least the mean of the quantum evolution gives us back some mean of the classical evolution. In the following, we shall have more occasions to reflect on a variety of other aspects of the physical interpretation of the formalism.

10.3 The Stationary State Equation and Its Resolvent

Any state $\psi(x, t)$ with a well-defined energy E is eigenstate of the energy operator $\hat{E}$

$$\hat{E}\psi(x, t) = E\psi(x, t), \tag{10.14}$$

so its time dependence is through the phase factor $e^{-i\frac{Et}{\hbar}}$, that is, $\psi(x, t) = \psi(x, 0)e^{-i\frac{Et}{\hbar}}$. So, the stationary states are solutions of

$$\left[\frac{\hat{p}^2}{2m} + V(\hat{x}, t)\right]\psi(x, t) = E\psi(x, t), \tag{10.15}$$

where E is the eigenvalue.

10.3.1 Hard Potential Wall

In one space dimension, one readily solves the problem with a potential wall $V(x) = \infty,\ x > 0, V(x) = 0,\ x < 0$, with the condition $\psi_E(x,t) = 0,\quad x < 0$;

$$\psi_E(x,t) \approx \sin(kx) \exp\left[\frac{-i\,E(p)t}{\hbar}\right],\ p = \hbar k. \tag{10.16}$$

This wave function represents an electron beam fired from a gun in the far left toward a potential wall. The particles bounce, and there is an incident wave[3] with momentum p and a reflected wave with momentum $-p$. By the interference between incident and reflected waves, the wave function vanishes for x multiple of $\frac{\pi}{k}$. This is not compatible with our classical intuition. The particles must pass through such points, but cannot be detected there, due to negative interference! The particles propagate as waves, and no corpuscles then exist; when the particles are detected somewhere the wave does not exist any more.

10.3.2 Bound States and Continuum States

Some potentials have no bound states, some have a finite number of bound states, while others have a numerable infinity of them, so that the generic nth eigenstate of H can be denoted more simply as $\psi_n(x,t) = \psi_n(x)e^{-i\frac{E_n t}{\hbar}}$. The probability density of such a state is time-independent. Therefore the energy eigenstates are called stationary states.

It can be shown that the stationary state with the lowest energy, usually called the ground state and denoted with $n = 0$, has no nodes (in other terms,$\psi_0(x) \neq 0$). Clearly, this also implies that the ground state cannot be degenerate (in other terms, there is only one ground state). Substituting

$$\psi_0(x,t) = \psi_0(x)e^{-i\frac{E_0 t}{\hbar}}, \tag{10.17}$$

into (10.7), we separate the variables. The time-independent equation for $\psi(x)$ reads as

$$\hat{H}\psi_0(x) = E_0\psi_0(x). \tag{10.18}$$

[3] By the Euler formula, $\sin(\alpha) = \frac{e^{i\alpha} - e^{-i\alpha}}{2i}$.

10.3.3 Resolvent

For a particle which is known to be at x at time 0, the probability amplitude to find it at y at time t is $\langle y|\hat{U}(t)|x\rangle$, where $\hat{U}(t)$ is the evolution operator defined in the last section. The Fourier transform $\langle y|\frac{1}{\omega-\hat{H}}|x\rangle$ can be phrased as the amplitude to go from x to y at frequency ω, and the resolvent

$$\hat{G}(\omega) = \frac{1}{\omega - \hat{H}}, \tag{10.19}$$

which is closely related to Green's functions, will be useful later.

10.4 Continuity Equation

The Continuity Equation is well known from Electromagnetism, where it states the conservation of charge. The charge density $\rho(\mathbf{x}, t)$ and the current density $\overrightarrow{J}(\mathbf{x}, t) = \rho\overrightarrow{v}$ (where $\overrightarrow{v}$ is the velocity) obey

$$\frac{\partial\rho}{\partial t} + \nabla\overrightarrow{J} = 0. \tag{10.20}$$

Integrating over volume V and using the Gauss Theorem,

$$\frac{d}{dt}\int_V \rho d^3x + \int_V \nabla\overrightarrow{J}\, d^3x = \frac{dQ}{dt} + \int_S \overrightarrow{J}\, d\overrightarrow{S} = 0. \tag{10.21}$$

The change total charge Q in a given volume depends on the flux of current across the boundary. In the single-particle Schrödinger theory, the relevant probability density is

$$\rho(\mathbf{x}, t) = |\psi(x, t)|^2,$$

and we must find the current density $\mathbf{J}(\mathbf{x}, t)$ associated with it. For charged particles, these expressions multiplied by the electron charge e naturally represent the charge and current densities associated with the wave function; again there is the need to average over many particles or many instances before allowing this correspondence.

In one dimension, the task of obtaining the current density is simplest; multiply by $\psi^*(x)$ the S.E.

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + V(x, t)\psi = \mathrm{i}\hbar\frac{\partial\psi}{\partial t},$$

then multiply by ψ the conjugated equation

$$-\frac{\hbar^2}{2m}\frac{d^2\psi^*}{dx^2}+V(x,t)\psi^*=-i\hbar\frac{\partial\psi^*}{\partial t}$$

and subtract: the result is

$$-\frac{\hbar}{2m}\left[\psi^*\frac{d^2\psi}{dx^2}-\psi\frac{d^2\psi^*}{dx^2}\right]=i\left[\psi^*\frac{\partial\psi}{\partial t}-\psi\frac{\partial\psi^*}{\partial t}\right], \tag{10.22}$$

that is,

$$-\frac{\hbar}{2mi}\frac{d}{dx}\left[\psi^*\frac{d\psi}{dx}-\psi\frac{d\psi^*}{dx}\right]=\frac{\partial|\psi|^2}{\partial t}.$$

Therefore, the probability current density is

$$J=\frac{\hbar}{2mi}\left[\psi^*\frac{d\psi}{dx}-\psi\frac{d\psi^*}{dx}\right]. \tag{10.23}$$

This result looks less unfamiliar if one reasons that $J=\frac{1}{2m}(\psi^* p\psi+c.c.)$. Since the probability of finding the particle becomes 1 if one looks for it everywhere, the wave function must be *normalized* to unity:

$$\int_{-\infty}^{\infty}dx|\psi(x,t)|^2=1. \tag{10.24}$$

This makes sense provided that the normalization is forever; we shall check that this is indeed the case provided that the potentials are real; in other words, the evolution is unitary.

The above is easily extended to three dimensions.

$$\mathbf{J}=\frac{\hbar}{2mi}\left[\psi^*\nabla\psi-\psi\nabla\psi^*\right]=\frac{1}{2m}\left[\psi^*\overrightarrow{p}\,\psi-\psi\overrightarrow{p}^{\,*}\psi^*\right]. \tag{10.25}$$

If ψ is real, the current vanishes. Complex wave functions are needed to describe transport phenomena. If we introduce a magnetic field by the known rule $\overrightarrow{p}\rightarrow\overrightarrow{p}-e\overrightarrow{A}$, a new contribution to the current is generated, namely, the diamagnetic current

$$\mathbf{J}_d=-\frac{e\mathbf{A}}{m}. \tag{10.26}$$

10.5 Schrödinger and Heisenberg Formulations of Quantum Mechanics

The *matrix mechanics* was formulated in 1925 by Werner Heisenberg.[4] In the same year the quantum theory of the Hydrogen atom was proposed by Wolfgang Pauli[5] in a beautiful but complicated applications of Group theory, assuming quantum commutation relations. In 1926 Erwin Schrödinger[6] wrote his fundamental paper mentioned above.

Initially, Schrödinger found the Heisenberg *matrix mechanics* unconvincing, while Heisenberg thought that Schrödinger's theory was rubbish. It did not take a long time before it became clear that the different formulations gave the same results because they were mathematically equivalent. Actually, each approach clarified the others. Both are equivalent to the Bohm theory to be discussed below and to Feynman's Path Integral formulation that will be presented in Sect. 15.7.

[4] Werner Heisenberg (Würzburg, 1901–München 1976), received Nobel prize in 1932.

[5] Wolfgang Pauli (Vienna 1900–Zurich 1958) received the Nobel prize in 1945.

[6] Erwin Schrödinger (Vienna 1887–Vienna 1961) succeeded to Max Planck as professor of Physics in Berlin in 1927; he won the Nobel Prize in 1933.

Chapter 11
Particle in One Dimension

Simple cases trigger Intuition.

11.1 Deep Potential Well

One-dimensional problems illustrate in simple fashion many aspects of the theory, and for this reason, every beginner in Quantum Mechanics is traditionally exposed to them. Some interesting one-particle problems in higher dimensions can be solved by separation of variables, thereby reducing them to one-dimensional subproblems, and we shall encounter several examples in other chapters. However, in recent times there have been important technological developments in nano-devices and low-dimensional objects (like nano-wires and nano-tubes) that have made these problems much more directly relevant. In classical physics, a wire is one-dimensional in the limited sense that its length is many times its diameter. However, in Quantum Mechanics, the low dimensionality of thin films and one-dimensional objects has a much stronger meaning. The next problem (deep potential well) shows that the confinement implies quantized levels divided by gaps that become wider the narrower the well. This implies that at low enough temperatures, the electrons behave as if the transverse dimensions were effectively frozen out. In this way, low-dimensional problems are no longer academic.

Consider a Schrödinger particle trapped between $x = 0$ and $x = a$ by the potential

$$V(x) = \begin{cases} \infty \text{ for } x < 0, \\ 0 \quad \text{for } 0 < x < a, \\ \infty \text{ for } x > a. \end{cases}$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_11

We must solve

$$\begin{cases} -\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi_n(x) = E_n\psi_n(x), \\ \psi_n(0) = \psi_n(a) = 0. \end{cases}$$

The result is

$$\psi_n(x) = C\sin(\chi_n x),\ \ \chi_n = n\frac{\pi}{a}, \tag{11.1}$$

where $n = 1, 2, \ldots$ is an integer quantum number (positive, since negative numbers give the same solutions). The energy eigenvalues are

$$E_n = \frac{\pi^2\hbar^2}{2ma^2}n^2. \tag{11.2}$$

Remarks:

1. Classically, the motion in this potential is too stupid to have the status of a problem, but the quantum version is very interesting. The energy spectrum is discrete. *Natura facit saltus!*
2. ψ_n has $n-1$ nodes (that is, $\psi_n(x) = 0$ has $n-1$ roots.) As n grows, λ gets shorter while E grows.
3. E_1 is the ground state, the others are excited states. Classically, the particle could be somewhere with zero velocity and zero energy. The minimum quantized energy is $E_1 > 0$; there is a minimum amount called *zero-point energy.* This is in line with the *uncertainty principle* $\Delta x\,\Delta p \sim \hbar$, where Δx is the spread in position and Δp is the spread in momentum. Since the particle is known to be between $x = 0$ and $x = a$, that is $\Delta x \sim a$, the momentum has uncertainty $\Delta p \sim \frac{\hbar}{a}$; therefore, $\Delta\frac{p^2}{2m} \sim E_1$.
4. $\psi_n(x)$ is real and there is no current. One can imagine that the particle bounces and creates two opposite currents that make up zero.

The normalization constant C is found by the condition

$$\int_0^a |\psi_n(x)|^2 dx = C^2\int_0^a dx\,\sin^2(\chi_n x) = 1.$$

Since

$$\int dx\,\sin^2(x) = \frac{x}{2} - \frac{\sin(2x)}{4},$$

one finds that the normalized eigenfunctions are

$$\psi_n(x) = \sqrt{\frac{2}{a}}\sin\left[n\pi\frac{x}{a}\right]. \tag{11.3}$$

Several remarks are in order:

1. $\psi_n(x)$ is the sum of opposite momentum plane waves. The mean momentum is

$$\langle p\rangle = \langle\psi_n(x)|p|\psi_n(x)\rangle = \int_0^a \psi_n(x)(-i\hbar)\frac{d}{dx}\psi_n(x)dx.$$

   This is imaginary. Are we in trouble? No, since $\langle p\rangle = 0$. Indeed, $\frac{1}{2}\int_0^a dx\frac{d}{dx}\psi_n(x)^2 = 0$.
2. $\psi_n(x)$ is eigenfunction of p^2 for x inside the well. The same argument applies to any bound state.
3. $\{\psi_n(x)\}$ is just the complete orthonormal set in the interval $(0, a)$ familiar from the theory of Fourier series.

11.2 Free Particle and Continuum States

It is time to worry about orthonormalization and physical meaning of the De Broglie waves, which represent the particle in a perfectly monochromatic, collimated beam; in one dimension,

$$\psi_k(x,t) = \frac{exp[i(kx - \omega(k)t)]}{\sqrt{2\pi}}. \tag{11.4}$$

11.2.1 Normalization

We encountered the factor $\frac{1}{\sqrt{2\pi}}$ when dealing with the De Broglie wave in (9.8), where it comes with the third power because in that case the problem is three-dimensional. It is there to ensure the correct mean value $\langle p\rangle$ of the momentum according to (9.11). The normalization method is different from the condition

$$1 = \int |\psi(x)|^2 dx,$$

which we use for bound states. Indeed, $\psi_k(x,t)$ is not $\in\in L^2$ (square integrable functions) over the x axis; we normalize it over the δ function:

$$\int_{-\infty}^{\infty} dx\psi_k(x,t)^*\psi_{k'}(x,t) = \delta(k-k'). \tag{11.5}$$

However, this is not the only method in use. One can always pretend that the electron gun and the entire experimental setup are contained in a box of side L; in one dimension, the plane wave function is

$$\psi_k(x,t) = \frac{exp[i(kx - \omega(k)t)]}{\sqrt{L}}. \tag{11.6}$$

We choose the discrete k values such that the wave functions are orthogonal over the Kronecker δ:

$$\delta_{m,n} = \begin{cases} 1, & m = n, \\ 0, & m \neq n. \end{cases}$$

This is granted by periodic boundary conditions

$$e^{ikL} = 1; \tag{11.7}$$

these are verified for a discrete set of k such that

$$k_n = \frac{2\pi n}{L}, \ n \text{ integer},$$

and the orthogonality is ensured by the theory of Fourier series. If L is large, the box has no influence on the experiment.

Dirac's δ yields a formal method for going to the infinite box limit. In three dimensions, the volume is $V = L^3$, and

$$\psi_k(x,t) = \frac{exp[i(kx - \omega(k)t)]}{\sqrt{V}}.$$

11.2.2 Phase Velocity

The dispersion law is the law $E(p)$, where E is the kinetic energy and p the momentum; it depends on the particle and also on the theory. For a corpuscle in classical mechanics, $E(p) = \frac{p^2}{2m}$, but in the relativistic theory, $E(p) = \sqrt{p^2c^2 + m^2c^4}$; for a photon, $\omega = ck$. More exotic, anisotropic dispersion laws describe the propagation of excitations (band electrons, dressed electrons, vibrations called phonons, mixed photon-vibration modes called polaritons) in condensed matter.

Although Schrödinger's equation is not the wave equation, nevertheless, the De Broglie plane wave is still of the form $f(x - v_p t)$, where v_p is called the phase velocity. For a photon it is the actual velocity c, but with the particle dispersion law $\hbar\omega(k) = \frac{p^2}{2m}$, the phase velocity $v_p = \hbar\frac{k}{2m}$ is half the physical speed $v = \frac{p}{m}$. Never mind, the wave function cannot be observed. In the plane wave the position is not defined, so the best one can do to measure the speed is to measure $\boldsymbol{p}$ and divide by m. The velocity operator is, indeed, $\boldsymbol{v} = \frac{\boldsymbol{p}}{m}$.

11.2.3 Group Velocity of a Wave Packet

Consider a free particle (potential $V = 0$). If we prepare a wave packet at time $t = 0$

$$\psi_0(x) \equiv \psi(x, 0) = \int dk \phi(k) e^{ikx}, \tag{11.8}$$

it will evolve at time t to become

$$\psi(x, t) = \int_{-\infty}^{\infty} dk \phi(k) e^{ikx} e^{-i\omega(k)t};$$

to see that, we could plug it into the S.E., but it is enough to observe that each k component in (11.8) is evolving as a De Broglie wave.

Suppose $\phi(k)$ is strongly peaked around $k = k_0$. This is a practically feasible approximation of a plane wave state. Only k values close to k_0 matter; in this neighborhood,

$$\omega(k) \sim \omega(k_0) + v(k - k_0), \quad v = \frac{d\omega(k)}{dk}|_{k_0}. \tag{11.9}$$

Therefore, to a good approximation,

$$\psi(x, t) = \int_{-\infty}^{\infty} dk \phi(k) e^{ikx} e^{-i[\omega(k_0) + v(k-k_0)\ldots]t},$$

that is,

$$\begin{aligned} \psi(x, t) &\sim \exp[i(-\omega(k_0)t + k_0 vt)] \int dk \phi(k) \exp[i(k[x - vt])] \\ &= \exp[i(-\omega(k_0)t + k_0 vt)] \psi_0(x - vt). \end{aligned}$$

The probability density $|\psi|^2$ propagates with the group velocity v, which is therefore the closest approximation to the classical behavior.

Continuing the expansion in (11.9), however, one discovers that the wave packet gradually broadens in real space.

11.3 Solutions of the Stationary State Equation in One Dimension

It follows from the S.E. that if the potential is continuous with all the derivatives (that is, $V(x) \in C^\infty$) the wave function $\psi(x)$ is also $\in C^\infty$. Depending on the form of the potential, one may find localized states, or running waves on the left or on the right. Localized states occur when $\psi(x) \to 0$ for $x \to \pm\infty$ and correspond to discrete

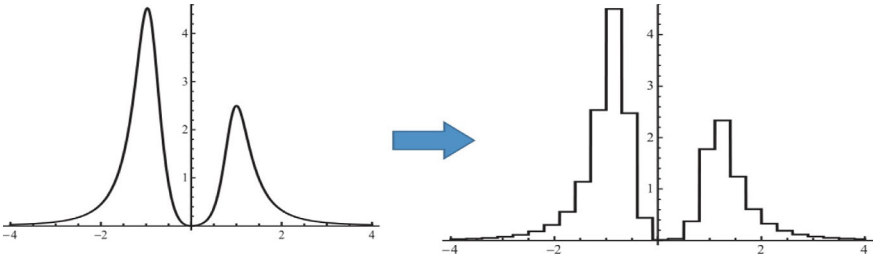


Fig. 11.1 A discretized potential introduces artificial steps, but makes the solution always easy, at least for the eigenfunctions of the low levels

energy eigenvalues for bound states. Many potentials of physical interest are scattering problems[1] in which for $x \to \pm\infty$ V tends toward constants and consequently ψ becomes a de Broglie wave. Then, all energies are possible, i.e., the spectrum of eigenvalues is continuous. In general, there are no closed-form analytical solutions. Thus, we can[2] approximate $V(x)$ by a piecewise constant, histogram-like function (see Fig. 11.1). This introduces artificial steps, but makes the solution easy. In any interval in which $V(x)$ is taken to be constant, $\psi(x)$ is a linear combination of two exponentials. In classically allowed regions where the energy E exceeds $V(x)$, these are $e^{\pm ikx}$, where $\hbar^2k^2 = 2m(E - V)$. Classically, the particle cannot be found where $E < V(x)$, but in those classically forbidden regions, the wave function continues; however, the exponentials become real, namely, $e^{\pm\lambda x}$ where $\hbar^2\lambda^2 = 2m(V - E)$. The need for a normalizable function forbids growing exponentials for $x \to \infty$ and decreasing exponentials for $x \to -\infty$.

At a step discontinuity of the potential at $x = x_0$, the S.E. implies a step-like contribution to $\psi''(x)$ proportional to $\theta(x - x_0)$, but $\psi'(x)$ and $\psi(x)$ remain continuous. In the case of the hard wall and at the ends of the deep well $\psi'(x)$ is discontinuous, but there V makes an infinite jump.

11.4 Potential Step

A particle in the potential

$$V(x) = \begin{cases} 0, & x < 0 \\ V > 0, & x > 0 \end{cases} \tag{11.10}$$

[1] There is indeed a variety of interesting problems with V, which is periodic (electrons in perfect crystals), aperiodic (e.g., electrons in glasses): here, I consider only the simplest examples.

[2] At high energies, the wave functions have high wave vector components and become sensitive to the discontinuities. However one is often interested in the low-lying states.

is free to go to infinity (both classically and in Quantum Mechanics); therefore, the spectrum of $\hat{H}$ is continuous. Suppose an electron gun fires a beam of particles from the left ($x = -\infty$) to a detector to the far right. We are looking for the stationary states, with a flux of particles reaching the barrier; classical bullets would bounce elastically if their kinetic energy E_k is less than the threshold V, otherwise they should go to the right with kinetic energy $E_k - V$.

11.4.1 Energy Below Threshold

If $E_k = \frac{\hbar^2 k^2}{2m} < V$, the solution is

$$\psi(x) = \begin{cases} Ae^{ikx} + Be^{-ikx} \text{ con } \ k > 0 \quad \text{per } x < 0, \\ Ce^{-\lambda x}, \ \text{per } x > 0, \end{cases} \tag{11.11}$$

with $\hbar^2 k^2 = 2mE$, $\hbar^2 \lambda^2 = 2m(V - E)$. The positive exponential would make the wave function blow up at infinity and was discarded.

Computing the current with the term in A, one finds that $J = |A|^2 \frac{p}{m} > 0$, where $p = \hbar k$; this is just the current from the gun, that is set by the experimenter; therefore we may take $A = 1$. The wave functions and its derivative are continuous:

$$\begin{cases} 1 + B = C, \\ ik(1 - B) = -\lambda C \Longrightarrow 1 - B = \frac{i\lambda}{k} C. \end{cases} \tag{11.12}$$

The sum of these equations yields $C = \frac{2k}{k+i\lambda}$; hence, $B = C - 1 = \frac{-i\lambda+k}{i\lambda+k} = -\frac{\lambda+ik}{\lambda-ik}$. Since $|B|^2 = 1$, B is a phase factor, and we may set $B \equiv e^{i\beta}$, with real β, while $C = 1 + e^{i\beta}$. Therefore, $C = e^{i\frac{\beta}{2}}(e^{-i\frac{\beta}{2}} + e^{i\frac{\beta}{2}})$ and the whole wave function becomes real if we multiply it by $e^{-i\frac{\beta}{2}}$. A constant phase factor is irrelevant, but it makes evident that this, like any real wave function, carries zero current. All the particles bounce. This is also evident from the fact that $\psi \to 0$ for $x \to \inf$: since $\frac{\partial \rho}{\partial t} = 0 \Rightarrow \frac{dJ}{dx} = 0$, $J = 0$ everywhere.

As in the infinite wall case, there are places where the particle cannot be found because of interference. On the other hand, there is a chance of detecting the particle in the classically forbidden region $x > 0$. The classical prohibition is absolute, since a corpuscle with $E = T + V$ and $V > E$ must have a negative kinetic energy, which is impossible. Even in Quantum Mechanics, the kinetic energy is a *positive definite operator* $\hat{T}$, that is, an operator such that all the eigenvalues are positive; any measurement of a positive definite operator must return a positive value. The paradox can be understood as follows. The classical concept that the particle at a given point has some kinetic energy is not tenable in Quantum Mechanics, since $\hat{T}$ does not commute with the position $\boldsymbol{x}$. We shall see in the next chapter that two observables exist simultaneously when the corresponding operators commute. The

kinetic energy matrix taken over a complete set has positive eigenvalues, but the x values in the barrier do not constitute a complete set, so the idea of a *negative kinetic energy under the barrier* is not sound. On the other hand, the particle propagates as a wave ψ with some energy E, and the corpuscle exists only after the detection. In the same way, in the double slit experiment, the wave goes through both slits, and the corpuscle exists only where the interference pattern is recorded. These remarks remove the paradox, but the quantum behavior remains qualitatively different!

11.4.2 Energy Above Threshold

If $E > V$,

$$\psi(x) = \begin{cases} e^{ikx} + Be^{-ikx}, & x < 0 \\ Ce^{i\chi x}, & x > 0, \end{cases}$$

with $\hbar^2 k^2 = 2mE$, $\hbar^2 \chi^2 = 2m(E - V.)$

From the continuity conditions for ψ and $\frac{d\psi}{dx}$ in $x = 0$

$$\begin{cases} 1 + B = C, \\ ik(1 - B) = i\chi C \Longrightarrow 1 - B = \frac{\chi}{k} C, \end{cases}$$

one finds that

$$\begin{cases} C = \frac{2k}{k+\chi}, \\ B = \frac{k-\chi}{k+\chi}. \end{cases}$$

The incident current is $A^2 k = k$; the reflected current is $|B|^2 k$; by definition, the ratio is the reflection coefficient

$$R = |B|^2 = \left| \frac{\chi - k}{\chi + k} \right|^2 .$$

Even if the barrier is lower than the energy eigenvalue E, a reflected current arises; this would not happen classically. The transmitted current is $|c|^2 \chi$; by definition, the transmission coefficient is the ratio with the incident current and is given by

$$T = |C|^2 \frac{\chi}{k} = 4\chi k \left| \frac{1}{\chi + k} \right|^2 .$$

Therefore, $R + T = 1$.

11.4.2.1 Descending Step $V < 0$

Classically, for a negative step ($V < 0$), all the particles should fall in with increased energy.

Quantum mechanically, the above solution still holds and there is a reflected current, too. This is observed experimentally when a beam of electrons is diffracted through the surface atoms of a solid.

11.4.3 Potential Well of Finite Depth

Consider the symmetric potential

$$V(x) = \begin{cases} 0, & x < -a \\ -W < 0, & -a < x < a \\ 0, & x > a. \end{cases}$$

An unsymmetric one would also be solvable, but would not add much to the interest. The symmetric case is special, but also instructive, because it offers an elementary example of the use of symmetry, which is a very important and far-reaching subject. An even potential ($V(-x) = V(x)$) allows for even solutions ($\psi(-x) = \psi(x)$) but also odd solutions ($\psi(-x) = -\psi(x)$).

Bound States $E < 0$

The bound solutions have negative energies, and there is a finite number of them; they come with two different symmetries:

even :

$$\psi(x) = \begin{cases} Be^{kx}, & x < -a, \\ D\cos(lx), & -a < x < a, \\ Be^{-kx}, & x > a, \end{cases}$$

odd :

$$\psi(x) = \begin{cases} -Be^{kx}, & x < -a, \\ D\sin(lx), & -a < x < a, \\ Be^{-kx}, & x > a, \end{cases} \tag{11.13}$$

where

$$k = \sqrt{\frac{-2mE}{\hbar^2}}, \qquad l = \sqrt{\frac{2m(E+W)}{\hbar^2}}. \tag{11.14}$$

Imposing the continuity conditions for ψ and $\frac{d\psi}{dx}$ in $x = a$, we are sure they are satisfied in $x = -a$ as well. Therefore (Fig. 11.2),

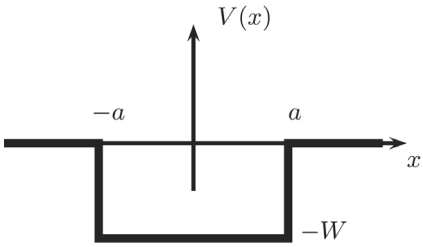


Fig. 11.2 Potential well of depth W

$$\begin{array}{ll} \text{even :} & \text{odd :} \\ \begin{cases} D\cos(la) = Be^{-ka}, \\ Dl\sin(la) = kBe^{-ka}; \end{cases} & \begin{cases} D\sin(la) = Be^{-ka}, \\ Dl\cos(la) = -kBe^{-ka}. \end{cases} \end{array} \tag{11.15}$$

The eigenvalue conditions are

$$\begin{array}{ll} \text{even :} & \text{odd :} \\ l\tan(la) = k & l\cot(la) = -k. \end{array} \tag{11.16}$$

Let us put

$$z = la = \sqrt{\frac{2m(E+W)}{\hbar^2}}a$$

in order to simplify the argument of transcendental functions. Equation (11.14) allows us to obtain k in terms of l. Using $l^2 = \frac{2mW}{\hbar^2} - k^2$,

$$k^2a^2 = \frac{2mW}{\hbar^2}a^2 - l^2a^2 = z_0^2 - z^2 > 0$$

with

$$z_0^2 = \frac{2mW}{\hbar^2}a^2;$$

so,

$$\frac{k}{l} = \frac{ka}{la} = \frac{\sqrt{z_0^2 - z^2}}{z}.$$

Equation (11.16) becomes

$$\begin{array}{ll} \text{even :} & \text{odd :} \\ \tan(z) = \frac{\sqrt{z_0^2 - z^2}}{z}, & \cot(z) = -\frac{\sqrt{z_0^2 - z^2}}{z}. \end{array} \tag{11.17}$$

These equations are readily solved numerically or graphically, and there are simple limiting cases.

If $z_0 \gg 1$, the well is deep, and if, in addition, we look at $z \ll z_0$, we obtain the lowest levels. As a first approximation, for $\zeta_0 \to \infty$,

$$\tan(z) \to \infty \; per \; z_n \to (2n+1)\tfrac{\pi}{2}, n \in \mathcal{N},$$
$$\cot(z) \to \infty \; per \; z_n \to (2n)\tfrac{\pi}{2}, n \in \mathcal{N},$$

where $\mathcal{N}$ is the set of natural numbers $1, 2, 3, \ldots$. The condition

$$z^2 = \frac{2m(E+W)a^2}{\hbar^2} = (2n+1)^2\frac{\pi^2}{4}$$

yields

$$E = -W + \frac{\hbar^2\pi^2}{2m(2a)^2}(2n+1)^2.$$

Taking into account that the width is $2a$, these are the odd n solutions of the deep well, which are spatially even; $\cot(z) \to \infty$ yields the even n solutions. When n grows this approximation deteriorates.

Since $0 < z < z_0$, if $z_0 \ll 1$, z must also be $\ll 1$. For spatially even states, we replace $z\tan(z) = \sqrt{z_0^2 - z^2}$ with $\tan(z) \approx z$, that is, $z^2 = \sqrt{z_0^2 - z^2} \Longrightarrow z^4 \approx z_0^2 - z^2 \Longrightarrow z^2 \approx z_0^2 - z_0^4$. There is always at least one solution. In the spatially odd case, we consider $z\cot(z) = -\sqrt{z_0^2 - z^2}$; there is no solution with $\cot(z) \approx \frac{1}{z}$. A very weak potential in one dimension always has a bound state, which is even.

The normalization requires, as usual, $\int_{-\infty}^{\infty} |\psi_n(x)|^2 = 1$.

11.4.4 Continuum States

For $E > 0$, all energies are allowed, and each eigenvalue is twice degenerate, that is, there are two orthogonal states with the same energy. This reflects the experimenter's freedom to choose between shooting with a gun from left or right. Let us choose the first arrangement. Classically, a marble thrown from $-\infty$ would proceed beyond the well with the same speed; because of quantum effects, part of the wave fired from $-\infty$ is reflected. Letting $\hbar k = \sqrt{2mE}$,

$$\psi(x) = \begin{cases} Ae^{ikx} + Be^{-ikx}, & x < -a, \\ C\sin(lx) + D\cos(lx), & -a < x < a, \\ Fe^{ikx}, & x > a, \end{cases}$$

and the incident amplitude is A, since e^{ikx} has the current $\frac{k}{m} > 0$ which goes to the right; B is the reflected amplitude and F the transmitted one. A is not really unknown, since the incident flux is chosen at will by the experimenter. The conditions in $-a$ yield

$$\begin{cases} Ae^{-ika} + Be^{ika} = -C\sin(la) + D\cos(la), \\ ik\left[Ae^{-ika} - Be^{ika}\right] = l\left[C\cos(la) + D\sin(la)\right], \end{cases} \tag{11.18}$$

and those in a give

$$\begin{cases} C\sin(la) + D\cos(la) = Fe^{ika}, \\ l\left[C\cos(la) - D\sin(la)\right] = ikFe^{ika}. \end{cases}$$

According to the Cramer[3] rule, since the determinant of coefficients is $-l$,

$$\begin{cases} C = Fe^{ika}\left[\sin(la) + i\frac{k}{l}\cos(la)\right], \\ D = Fe^{ika}\left[\cos(la) - i\frac{k}{l}\sin(la)\right]. \end{cases} \tag{11.19}$$

Solving (11.18), one finds

$$\begin{cases} A = \frac{e^{ika}}{2k}\left[-(Ck + iDl)\sin(la) + (Dk - iCl)\cos(la)\right], \\ B = \frac{e^{-ika}}{2k}\left[(-Ck + iDl)\sin(la) + (Dk + iCl)\cos(la)\right]; \end{cases} \tag{11.20}$$

and substituting (11.19),

$$\begin{cases} A = \frac{-ie^{2ika}}{2kl}F\left[(k^2 + l^2)\sin(2la) + 2ikl\cos(2la)\right], \\ B = \frac{-iF}{2kl}(k^2 - l^2)\sin(2la). \end{cases} \tag{11.21}$$

Finally,

$$F = \frac{e^{-2ika}}{\cos(2la) - i\frac{\sin(2la)}{2kl}(k^2 + l^2)}. \tag{11.22}$$

The reflection coefficient is $R = |\frac{B}{A}|^2$ and the transmission one is $T = |\frac{F}{A}|^2$; of course, $R + T = 1$.

[3] Gabriel Cramer (Geneve 1704–Bagnols sur Céze (France) 1752).

11.5 The δ Potential

The limit of a deep narrow potential

$$V(x) = -a\delta(x),\ a > 0 \tag{11.23}$$

is also interesting and lends itself to a simple calculation of the discrete spectrum. Very short range interactions also arise from the Weak Interaction. The S.E. for $x \neq 0$ reads as

$$\frac{d^2\psi}{dx^2} = k^2\psi(x), k^2 = -\frac{2mE}{\hbar^2} > 0.$$

The wave function is $\psi(x) = A[e^{kx}\theta(-x) + e^{-kx}\theta(x)]$. When doing the first derivative, the θ functions originate δ functions, which, however, cancel. When doing the second derivative, we get the δ function which is required by the Schrödinger equation. We find that the δ potential has one bound state, with

$$E = -\frac{ma^2}{2\hbar^2}.$$

The normalization constant A is easily obtained and

$$\psi(x) = \frac{\sqrt{ma}}{\hbar}\exp\left[-\frac{ma|x|}{\hbar^2}\right].$$

11.6 Potential Barrier and Tunneling

Consider a beam of particles fired from a gun, say, on the left, toward the potential bump

$$V(x) = \begin{cases} 0 & x < 0, \\ V > 0 & 0 < x < S, \\ 0 & x > S \end{cases}$$

with the detector on the right. When the beam energy $E > V$, that is, the kinetic energy is higher than the barrier, the problem is easily worked out, but the results are qualitatively similar to the $S \to \infty$ case that we have already considered above, with some reflected and some transmitted current. The famous tunnel effect arises when the energy is below threshold.

With $E < V$, the general integral is a combination of real exponentials $e^{\pm\lambda x}$ in the barrier, while outside, the exponentials are complex $e^{\pm ikx}$, $k > 0$. The $e^{\pm ikx}$ waves carry current $\pm k$. Since the gun is on the left, no particle comes from $x \to$ inf and the wave e^{-ikx} is absent for $x > S$. Using the rules of Sect. 11.3, the solution is found

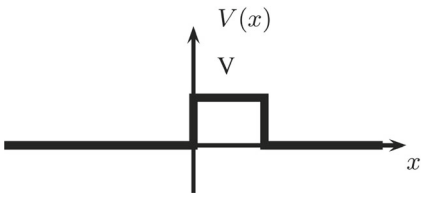


Fig. 11.3 Potential barrier

to be

$$\psi(x) = \begin{cases} \alpha e^{ikx} + \beta e^{-ikx} & x < 0, \\ \gamma e^{-\lambda x} + \delta e^{\lambda x} & 0 < x < S, \\ \epsilon e^{ikx} & x > S, \end{cases}$$

with $\hbar^2 k^2 = 2mE$, $\hbar^2 \lambda^2 = 2m(V - E)$. We may fix the incoming current by setting $\alpha = 1$. The continuity relations give us (Fig. 11.3):

$$\begin{cases} 1 + \beta = \gamma + \delta, \\ ik(1 - \beta) = -\lambda\gamma + \lambda\delta, \\ \gamma e^{-\lambda S} + \delta e^{\lambda S} = \epsilon e^{ikS}, \\ -\lambda\gamma e^{-\lambda S} + \lambda\delta e^{\lambda S} = ik\epsilon e^{ikS}. \end{cases} \tag{11.24}$$

The last two give us γ and δ: letting

$$\eta = \frac{ik}{\lambda},$$

one finds that

$$\begin{cases} \gamma = \frac{\epsilon e^{ikS}}{2}(1 - \eta)e^{\lambda S}, \\ \delta = \frac{\epsilon e^{ikS}}{2}(1 + \eta)e^{-\lambda S}. \end{cases}$$

The first two (11.24) give us

$$2 = \gamma\left(1 - \frac{1}{\eta}\right) + \delta\left(1 + \frac{1}{\eta}\right).$$

Hence,

$$\begin{aligned} 2 &= \tfrac{\epsilon e^{ikS}}{2}\left\{(1 - \eta)\, e^{\lambda S}\left(1 - \tfrac{1}{\eta}\right) + (1 + \eta)\, e^{-\lambda S}\left(1 + \tfrac{1}{\eta}\right)\right\} \\ &= \tfrac{\epsilon e^{ikS}}{2}\left\{2\cosh(\lambda S) - \left(\eta + \tfrac{1}{\eta}\right)\sinh(\lambda S),\right\} \end{aligned}$$

and finally,

$$\epsilon e^{ikS} = \frac{1}{\cosh(\lambda S) - \frac{1}{2}\left(\eta + \frac{1}{\eta}\right)\sinh(\lambda S)}.$$

The classically impenetrable barrier is overcome with probability $|\epsilon e^{ikS}|^2 \sim e^{-2\lambda S}$. This is the famous **tunnel effect**. In Sect. 11.4.1 I have already discussed the paradox of a particle that is detected in a classically forbidden region; the tunnel effect is an even more striking piece of evidence, since one can measure the energy E of the particle beyond the barrier even at a long distance from it, find that $E < V$ and reach the conclusion that the particle has crossed a forbidden region. This can be understood only if the particle propagates as a wave ψ with some energy E and the corpuscle exists as such only after the detection in some position. If a detector detects the particle outside the barrier, there is no paradox; if it were to detect the particle while in the barrier, the detection process would also have to supply the energy needed for the particle to be there.

There are countless examples and applications of tunnel phenomena in Science and Technology. Here is a short list:

1. Nuclear decay. A rough idea of α and β decays is provided by a potential well with a transparent wall (Fig. 11.4). One can start by evaluating the bound states ϕ_n that the well would have if the wall were infinitely thick. When the actual potential is restored, ϕ_n is no longer an eigenstate, but rather a wave packet called a *resonance*, with a width ΔE in energy. Sooner or later the particle will escape to infinity. The interested reader can find more about resonances in Chap. 24. The life time Δt of the resonance can vary from a fraction of a second to several billion years.
2. Hydrogen bond. In liquid H_2O, a proton can tunnel between two Oxygen atoms like a particle in a double well potential. This gives a significant contribution to the chemistry of water and other liquids with OH groups. Normally isotope effects are negligible, but beware! Heavy water D_2O with Deuterium replacing Hydogen is poisonous!
3. Josephson effect. In superconductors, electrons form bound pairs called Cooper pairs (see Sect. 25.6.1). This happens at low temperatures when the Coulomb repulsion is exceeded by an effective attractive interaction due to vibrations, or

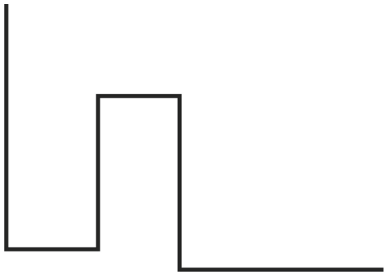


Fig. 11.4 Potential well with a penetrable wall. In 1928, George Gamow used this as a simple model for the nuclear alpha decay

to special electron correlation effects. If two superconductors separated by a thin insulating barrier are placed in a circuit, Josephson discovered that the application of a d.c. potential difference V produces an alternating current

$$I = I_0 \sin(\frac{2eV}{\hbar}t),$$

while under some conditions, an a.c. $V(t)$ produces a d.c. current (Shapiro effect). Such interesting phenomena can be understood in terms of Cooper pairs tunneling the barrier as such. Their effective wave function $\Psi(\boldsymbol{x}, t)$ is called the superconducting order parameter. An account of superconductivity is in Chap. 26.

There are also many applications. For example:

1. SQUID (Superconducting Quantum Interference Devices) are extremely sensitive magnetometers. They can measure 10^{-14} T, while for instance the magnetic fields produced by the heart are of order 10^{-10} T and those produced by the brain can also be measured. The SQUID is based on the Josephson effect.
2. In nanocircuits, quantum effects are important.
   The metal–insulator–metal junction devices, and the MOSFETs (metal oxide semiconductor field effect transistor), which are currently widely used in electronics, use an electric field to modulate the tunnel effect through a thin insulating barrier to change the electron population and conductivity.
3. The single-electron transistor Fig. 11.5 is also based on the tunnel effect. The gate potential modulates the barrier and therefore also the current produced by the potential V.
4. The tunnel effect microscope Fig. 11.6 allows us to scan and image a surface with atomic resolution with a metal tip. It was invented in 1981 by IBM scientists Gerd Binnig e Heinrich Rohrer, who won the Nobel for Physics in 1986. While the tip scans the surface, its distance from it, that is of the order of atomic sizes,

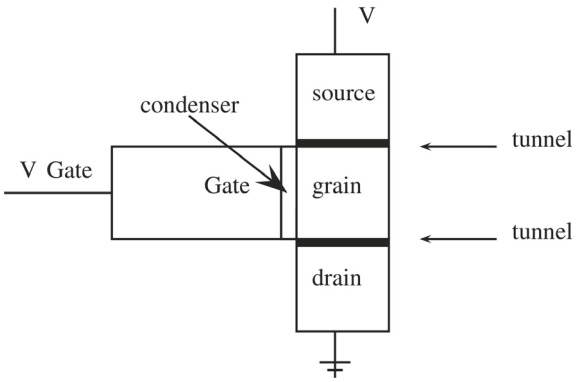


Fig. 11.5 SET (Single-electron tunneling) is a transistor. The grain is a metallic island of nanoscopic size (10^{-6} cm or even less.)

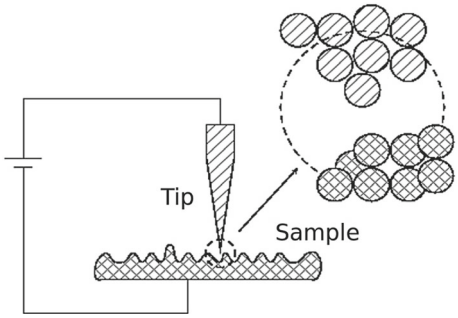


Fig. 11.6 Block diagram of the tunnel microscope. In the enlarged detail, it is shown that the tip ends in a single atom from which the current goes to the surface

is recorded. The precise positioning of the tip is achieved by the piezoelectric effect. This effect arises in some solids like SiO_2 with a low-symmetry ionic crystalline structure that produces an electrical potential if deformed. Very sensitive transducers are built on this principle. A potential difference between the tip and the surface drives a tunnel current: the vacuum acts as barrier. When the tip is positive, the current comes from filled electronic states, when it is negative electrons are injected into empty states of the sample. This current depends exponentially on the distance from the surface. In this way, one gains information about morphology and electronic structure.

Tunneling time

Since the early days of quantum mechanics various estimates of the time needed for a particle to tunnel a barrier have been published. This is not a simple problem. While it is clear that the non-relativistic theory does not consider the limit of the velocity of light, it is also clear that there is a broad distribution of tunneling times, particularly when the tunneling probability is small. A 2014 paper by Alexandra S. Landsman and coworkers on the journal optica presents experiments of laser ionization of He atoms and compares the various approaches. Alexandra S. Landsman et al., Vol. 1, No. 5/November 2014/Optica 343.

Chapter 12
The Postulates of Quantum Mechanics: Postulate 1

The axiomatic formulation of Quantum Mechanics is generally presented as a set of four postulates, introduced by John von Neumann. The physical meaning of each of them requires a nontrivial, careful enquiry.

Here is the first postulate.

> The state of any system is represented by a complex wave function $\Psi_a(x, t)$, where x stands for the set of all the coordinates, t denotes time, and a is a (possibly empty) set of constants of the motion (the so-called quantum numbers). If a is the set of all the observables that are compatible, the quantum state is uniquely determined.

12.1 The Wave Function

The system referred to above could be a particle, an atom, or even a macroscopic superconductor or a neutron star (then, x stands for a very large set of coordinates), so the statement is quite strong and general. In any case, all the information that is available from all possible experiments is in $\Psi_a(x, t)$. The wave function must be taken to be *normalized* such that the probabilities sum to unity. In the case of a single degree of freedom $\Psi_a(x, t)$, the normalization condition reads as $\int |\Psi(x, t)|^2 dx = 1$, while in general, one must integrate the square modulus over all the variables. The function is complex, therefore $\Psi(x, t) = |\Psi(x, t)|e^{i\phi(x,t)}$, where $\phi(x, t)$ is the phase. One can

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_12

change $\phi(x,t)$ by a constant phase factor (for instance, multiplying $\Psi(x,t)$ by the imaginary unit i), but the physical state remains the same; nevertheless, the phase difference between two wave functions does matter a lot, since the wave functions do interfere. We shall see that the phase can be changed in several ways (e.g., rotations, Galileo transformations and gauge changes) and brings physical information in a nontrivial way (e.g., through the Berry phase, to be discussed below).

The set of possible wave functions of a given system is a complex vector space; all linear combinations of wave functions representing alternative states of the same system $\Phi = \alpha\Psi + \beta\chi$ with complex coefficients α, β, once normalized, are other possible wave functions. The scalar product defined as

$$(\Psi|\Phi) \equiv \langle\Psi|\Phi\rangle = \int \Psi(x,t)^* \Phi(x,t)\, \mathrm{dx} \tag{12.1}$$

is a complex number. In this equation two alternative notations, both in use, are presented. The above notation $\langle\Psi|\Phi\rangle$ is the bra–ket notation introduced by Dirac, with $\Psi(x)$ as a bra and $\Phi(x)$ as a ket. When $\langle\Psi|\Phi\rangle = 0$, the functions are said to be orthogonal. The maximum number d of orthogonal wave functions is the dimension of the vector space, for many systems $d = \infty$. The normalization condition is $\langle\Psi|\Psi\rangle = 1$. Most often, it is preferable to work with an orthonormal basis set of functions. Let us see how to achieve such a set.

12.2 Gram–Schmidt Orthogonalization

The Gram–Schmidt Orthogonalization is the most obvious. Let $(v_1, \ldots, v_n)$ be a set of normalized vectors, taken in any order. Using any orthonormal basis, $(v_i|v_j) = \sum_\alpha v_{i,\alpha}^* v_{j,\alpha}$ denotes the scalar product. We choose $w_1 = v_1$ as the first vector of the orthogonalized set and remove the component along v_1 from the remaining vectors by $v_i \to v_i - (v_i|v_1)v_1$. We normalize all the new vectors, then set $w_2 = v_2$ as the second vector of the orthogonalized set; next we proceed to orthogonalize the vectors $(v_3 \cdots v_n)$ to w_2, and so on. The procedure fails to give the basis if the initial set does not span an n-dimensional space. The result of this procedure is a basis of orthogonal vectors, but if we change the order of the v_i we get a different basis.

12.3 Löwdin Symmetric Orthogonalization

A smart orthogonalization method, proposed by Löwdin in 1950, is often preferred for its aesthetic and also practical advantages. It is called symmetric orthogonalization, since it starts from a set of normalized non-orthogonal vectors $(v_1, \ldots, v_n)$ and treats all of them on equal footing.

To start with, one computes the overlap matrix S with entries

$$S_{mk} = (v_k|v_m). \tag{12.2}$$

We note that S is Hermitian ($S^*_{km} = S_{mk}$). Then, one looks for a set of coefficients L_{ij} of a linear transformation

$$w_i = \sum_{j=1}^{n} L_{ij} v_j \tag{12.3}$$

such that the new basis is orthogonal, that is

$$(w_i|w_j) = \left(\sum_{p=1}^{n} L_{ip} v_p | \sum_{q=1}^{n} L_{jq} v_q \right) = \sum_{p,q=1}^{n} L^*_{ip} L_{jq} S_{qp} = \delta_{ij}. \tag{12.4}$$

The matrix L is our unknown here, and it is far from unique, since there are infinite possible different sets of orthogonal vectors w_i. We can exploit this arbitrariness to put conditions on L. It is convenient to pick a Hermitian L matrix (that is, $L^*_{ik} = L_{ki}$), then, the condition becomes

$$(w_i|w_j) = \sum_{p,q=1}^{n} L_{jq} S_{qp} L_{pi} = \delta_{ij}. \tag{12.5}$$

We can rewrite the condition (12.5) in terms of matrices as

$$LSL = 1, \tag{12.6}$$

where 1 obviously stands for the unit matrix. A solution is now evident:

$$L = S^{-\frac{1}{2}}. \tag{12.7}$$

So, let us look for a Hermitian L. This is not hard to do. Since the overlap matrix S is Hermitian, it can be diagonalized: $S_d = U^\dagger S U = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, where the columns of U are the eigenvectors of S and λ_i is the ith eigenvalue. Then, we can denote by the symbol $S_d^{-\frac{1}{2}}$ the diagonal matrix

$$S_d^{-\frac{1}{2}} = \mathrm{diag}\left(\lambda_1^{-\frac{1}{2}}, \lambda_2^{-\frac{1}{2}}, \ldots, \lambda_n^{-\frac{1}{2}} \right). \tag{12.8}$$

Hence,

$$S^{-\frac{1}{2}} = U S_d^{-\frac{1}{2}} U^\dagger = L \tag{12.9}$$

does the job. Here is the desired transformation to the symmetric orthonormal set:

$$w_j = \sum_{k=1}^{n} \left(S^{-\frac{1}{2}} \right)_{kj} v_k. \tag{12.10}$$

Besides being convenient for numerical work, this method has been shown to minimize $\sum_i^n |w(i) - v(i)|^2$; in this way, the orthogonal basis is as close as possible to the original one.

12.4 Schwarz Inequality

Let us first consider a real vector space and let $\overrightarrow{a}$, $\overrightarrow{b}$ be two vectors. The norms are $a = \sqrt{\overrightarrow{a} \cdot \overrightarrow{a}}$ and $b = \sqrt{\overrightarrow{b} \cdot \overrightarrow{b}}$.

Then, $\overrightarrow{a} \cdot \overrightarrow{b} = ab\cos(\theta)$, where θ is the angle between $\overrightarrow{a}$ and $\overrightarrow{b}$; since the square of the cosine of any real angle is ≤ 1, one finds the *Schwarz inequality*

$$|\overrightarrow{a} \cdot \overrightarrow{b}|^2 \leq (\overrightarrow{a} \cdot \overrightarrow{a})(\overrightarrow{b} \cdot \overrightarrow{b}).$$

This is elementary, but the interesting thing is that it holds for any real or complex vector space even with infinitely many dimensions. Therefore, even for state vectors (that is, wave functions),

$$|\langle \alpha | \beta \rangle|^2 \leq \langle \alpha | \alpha \rangle \ \langle \beta | \beta \rangle = 1.$$

12.5 Hilbert Spaces

Every state of the system is represented by a normalized function $\Psi_a(x, t)$; the function is a normalized vector $|a\rangle$ in a possibly infinite space of functions. The existence of a norm requires the notion of the length of a vector, or, equivalently, the notion of the distance between two points. So, the space is a normed space. The expansion of a vector over an infinite basis is a series expansion and requires a limit to exist; physically, we must require that each converging series represents a possible state. Therefore, we must ask that the limit of every convergent series belongs to the space. However, this is not granted in an infinite-dimensional normed vector space. For instance, the Taylor series for $\sin(x)$ is a polynomial series, but $\sin(x)$ is not a polynomial, so the space of polynomials fails to satisfy this requirement. When every convergent series belongs to the space, we speak of a complete space. A complete normed space is called a Banach space. However there are many ways

to assign a norm in a Banach space. For example, the space of continuous functions $f(x)$ defined in a closed interval (for instance, $x \in [0, 1]$) is a Banach space with the norm $||f|| = sup|f(x)|$. The space of square integrable functions where the wave functions belong is called L^2. The completeness of L^2 is shown by the Fourier theorem. It is a special Banach space in which the norm is defined in terms of the scalar product; in one dimension, it is $||f(x)|| = \langle f|f\rangle = \int_{-\infty}^{\infty} |f(x)|^2$; this definition extends to many dimensions in the obvious way. Such a Banach space is called a Hilbert space.

Suppose a system S_1 has a Hilbert space H_1 of dimension d_1 with basis $a_1, a_2, \ldots a_{d1}$, and a second system S_2 has a Hilbert space H_2 of dimension d_2 with basis $b_1, b_2, \ldots b_{d2}$. Now, the two systems might interact. To handle the compound system $S_1 \bigcup S_2$, we need the tensor product $H_1 \bigotimes H_2$ with basis $a_1 \bigotimes b_1, a_1 \bigotimes b_2, \ldots a_{d1} \bigotimes b_{d2}$.

12.6 Quantum Numbers

Depending on the way a system is prepared, the wave function Ψ may be an eigenfunction of certain operators. Then, it is very helpful to label Ψ as $\Psi_a(x)$, where a stands for the set of eigenvalues of those operators. This might be an empty set, otherwise $a = \{A, B, C, \ldots\}$ and its entries are called quantum numbers. We shall see that the labels must correspond to compatible observables, and their operators must commute. In the case in which Ψ is an eigenfunction of a time-independent Hamiltonian $\hat{H}$, the set $a = \{A, B, C, \ldots\}$ comprises the energy and constants of the motion that coexist; they are referred to as compatible observables; they must commute both with $\hat{H}$ and each other. For a given $\hat{H}$, there are, at times, alternative choices of the set of operators, and the label a is a subset of one of those. For the simple case of a free Schrödinger particle, as we shall see, the state can be labeled by energy and 3 momentum components, or by energy, the square of angular momentum and one of its components, but the square of angular momentum is not compatible with the components of the momentum, and the 3 components of the angular momentum are not compatible.

If a contains all the operators of a set of compatible operators, the quantum state is uniquely identified; this implies that two different quantum states must differ by at least one quantum number.

Chapter 13
Postulate 2

Any observable Q corresponds to an operator $\hat{Q}$ acting on the wave functions. These operators are linear, Hermitian (that is, $\hat{Q}^\dagger = \hat{Q}$), and must possess a complete set of eigenvectors. The expectation value of Q in the state $|\Psi\rangle$ is $\langle\Psi|\hat{Q}|\Psi\rangle \equiv \langle\Psi|\hat{Q}\Psi\rangle$.

13.1 Mathematical Context: Operators, Matrices and Function Spaces

$D \equiv \frac{d}{dx}$, and also $\hat{x}$ (which multiplies by x), are examples of linear operators $\hat{O}$: $\hat{O}(\Phi + \Psi) = \hat{O}\Phi + \hat{O}\Psi$. Coordinates and momenta of Classical Mechanics wear a hat and become quantum operators. Actually, the Hamiltonian and all the other functions familiar from Classical Mechanics are also important in Quantum Mechanics, but become quantum operators; not only that, but we shall meet more of them, having no classical counterpart, like spin, particle exchange and more. The scalar product of $\hat{Q}|\Psi\rangle$ with $|\Phi\rangle$, which is $\langle\Phi|\hat{Q}|\Psi\rangle$, may be regarded as the element Q_{ij} of some matrix Q with i and j replaced by the indices Φ and Ψ, which are functions $\in L^2$, that is, their square modulus has a finite integral over all space; it is called a matrix element of $\hat{Q}$; the only real difference with a conventional matrix M_{ij} is that the indices are often continuous and the matrix most often has an infinity of rows and columns. Werner Heisenberg initially formulated the theory in terms of matrices.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_13

13.2 Why Hermitian Operators?

Let α, β denote vectors of the Hilbert space. The Hermitian adjoint $\hat{A}^\dagger$ of $\hat{A}$ is defined by

$$\langle \hat{A}^\dagger \alpha | \beta \rangle = \langle \alpha | \hat{A} \beta \rangle \quad \forall \alpha, \beta. \tag{13.1}$$

In the one-dimensional case, the notation is simplest: then, we write

$$\int dx (\hat{A}^\dagger \alpha)^* \beta = \int dx \alpha^* \hat{A} \beta,$$

and taking the complex conjugate of both sides,

$$\left(\int dx \alpha^* \hat{A} \beta \right)^* = \int dx \beta^* \hat{A}^\dagger \alpha;$$

this suggests the general rule for taking the complex conjugate of any matrix element:

$$\langle \alpha | \hat{A} | \beta \rangle^* = \langle \beta | \hat{A}^\dagger | \alpha \rangle \quad \forall \alpha, \beta. \tag{13.2}$$

Self-adjoint or Hermitian operators $\hat{Q}$ have a special importance. A Hermitian operator is such that $\hat{Q}^\dagger = \hat{Q}$; hence all Hermitian operators have the property that $\langle \Phi | \hat{Q} \Psi \rangle = \langle \hat{Q} \Phi | \Psi \rangle$. If we set in Eq. (13.2) $\alpha = \beta$, we conclude that all the expectation values of Hermitian operators must be real. Since measured quantities must be real, the reader can have the impression that observables correspond to Hermitian operators. We shall see that this is true, and that the eigenvalues are also real. This is not the case for $D = \frac{d}{dx}$; in fact,

$$\langle f | Dg \rangle = -\langle Df | g \rangle.$$

and D is *anti-Hermitian*, while the momentum $\hat{p} = -i D$ is Hermitian[1] over L^2.

13.3 Orthogonal Spaces

Different eigenvalues of any Hermitian operator belong to orthogonal eigenvectors. Indeed, starting with

$$\begin{cases} \hat{A}|m\rangle = a_m |m\rangle, \\ \hat{A}|n\rangle = a_n |n\rangle, \end{cases}$$

[1] $\hat{p}$ fails to be Hermitian on the space $P(N)$ of polynomials of degree N, but this does not matter because all polynomials are outside L^2.

and taking scalar products, we get

$$\begin{cases} \langle n|\hat{A}|m\rangle = a_m\langle n|m\rangle, \\ \langle m|\hat{A}|n\rangle = a_n\langle m|n\rangle. \end{cases}$$

The complex conjugate of the second is $\langle n|\hat{A}^\dagger|m\rangle = a_n^*\langle n|m\rangle$, and since $A = A^\dagger$, it follows that $\langle n|\hat{A}|m\rangle = a_n\langle n|m\rangle$. Subtracting from the first equation, we find that $0 = (a_n - a_m)\langle n|m\rangle$. So, if $(t_n - t_m) \neq 0$, it follows that $\langle n|m\rangle = 0$.

13.4 Completeness

The eigenvectors of an observable operator $\hat{T}$ represent states such that $\hat{T}$ has a sharp value. The set of eigenvectors must be complete; this point will be expanded when dealing with the third postulate. The completeness grants that every $|\Psi\rangle$ can be expanded:

$$|\Psi\rangle = \sum_m |m\rangle\langle m|\Psi\rangle. \tag{13.3}$$

Here, $\sum_m$ is a sum where the spectrum of $\hat{T}$ is discrete and an integral where it is continuous. In general, $|\Psi\rangle$ and $|m\rangle$ depend on many variables, but in one-dimensional problems, this expansion is just the familiar Fourier theory:

$$f(x) = \int_{-\inf}^{\inf} dq \frac{e^{-iqx}}{2\pi} \tilde{f}(q).$$

This is the continuous expansion in the complete set of plane waves over the real axis, which would be replaced by a Fourier series in the case of a periodic system. The plane waves have sharp energy and momentum. Now we can see that if $\hat{T}$ is an observable, it is *physically necessary* that the set $\{m\}$ of eigenvectors be complete. The need for completeness arises because the knowledge of $f(q)$ must be equivalent to that of $f(x)$, since this is a canonical transformation already in Classical Mechanics.

Formally, we may rewrite (13.3) in the form

$$\mathbf{1} = \sum_m |m\rangle\langle m|, \tag{13.4}$$

where $\mathbf{1}$ is the identity operator. Therefore, if $A_{ij} = \langle\Phi_i|\hat{A}\Phi_j\rangle \equiv \langle\Phi_i|\hat{A}|\Phi_j\rangle$ is the matrix element of $\hat{A}$ between two elements of a basis set, and B_{ij} is the matrix element of $\hat{B}$, the product $\hat{A}\hat{B}$ turns out to be $\langle\Phi_i|\hat{A}\hat{B}|\Phi_j\rangle = \langle\Phi_i|\hat{A}\sum_k |k\rangle\langle k|\hat{B}|\Phi_j\rangle = \sum_k A_{ik}B_{kj}$ and *the matrices of the operators must by multiplied row by column.* Mathematically, the matrices are a *representation* of the operators. The operator is known if the matrix and the basis set are specified.

13.5 Commutators: Angle and Angular Momentum

Fanfare is needed here, because we are about to meet for the first time Planck's constant h= 6.626×10^{-34} Js, dimensionally an action. The whole quantum world exists because of her. Recall that, classically, the Poisson bracket of canonically conjugated variables is 1. In Quantum theory, instead, couples of canonically conjugated operators have commutators $-i\hbar$, where $\hbar = \frac{h}{2\pi}$. For instance,

$$[\hat{p}, \hat{x}] = \hat{p}\hat{x} - \hat{x}\hat{p} = -i\hbar. \tag{13.5}$$

Please do not ask "why?". Science aims to explain the observed phenomena in terms of fundamental facts, and this one is fundamental.

Now we meet another remarkable pair: angular momentum and angle. We can start from the example of the *plane rigid rotator*, which can rotate around the z axis, with moment of inertia[2] I. In terms of the angle ϕ, the Lagrangian is

$$L(\phi, \dot{\phi}) = \frac{1}{2} I \dot{\phi}^2.$$

The momentum $\hat{L}_z = I\dot{\phi}$ is canonically conjugated to the angle ϕ. The Hamiltonian operator is like the classical function:

$$\hat{H} = \frac{L_z^2}{2I}, \tag{13.6}$$

where $\hat{L}_z$ is the z component, $\hat{L}_z = xp_y - yp_x$.

In plane polar coordinates,

$$x = \rho \cos\phi, \;\; y = \rho \sin\phi.$$

The components of the momentum $-i\hbar\nabla$ can be obtained from the *chain rule*

$$\frac{\partial}{\partial x} = \frac{\partial \rho}{\partial x}\frac{\partial}{\partial \rho} + \frac{\partial \phi}{\partial x}\frac{\partial}{\partial \phi}, \quad \frac{\partial}{\partial y} = \frac{\partial \rho}{\partial y}\frac{\partial}{\partial \rho} + \frac{\partial \phi}{\partial y}\frac{\partial}{\partial \phi}.$$

So,

$$\rho = \sqrt{x^2 + y^2}, \;\; \phi = \arctan\frac{y}{x}.$$

Differentiating the radius,

$$\frac{\partial \rho}{\partial x} = \frac{x}{\rho} = cos\phi, \;\; \frac{\partial \rho}{\partial y} = \frac{y}{\rho} = sin\phi,$$

[2] For a body with density $\rho(\vec{x})$, the moment of inertia relative to the z axis is $I = \int d^3x \rho(\vec{x})(x^2 + y^2)$.

and since

$$\frac{d \arctan u}{du} = \frac{1}{1+u^2},$$

we obtain $\frac{\partial \phi}{\partial x} = \frac{1}{1+\left(\frac{y}{x}\right)^2} \frac{d}{dx}\left(\frac{y}{x}\right) = -\frac{y}{x^2+y^2}$, $\frac{\partial \phi}{\partial y} = \frac{1}{1+\left(\frac{y}{x}\right)^2} \frac{d}{dy}\left(\frac{y}{x}\right) = \frac{x}{x^2+y^2}$; in polar coordinates,

$$\frac{\partial \phi}{\partial x} = -\frac{sin\phi}{\rho}, \ \frac{\partial \phi}{\partial y} = \frac{cos\phi}{\rho};$$

so,

$$\frac{\partial}{\partial x} = cos\phi \frac{\partial}{\partial \rho} - \frac{sin\phi}{\rho} \frac{\partial}{\partial \phi}, \ \frac{\partial}{\partial y} = sin\phi \frac{\partial}{\partial \rho} + \frac{cos\phi}{\rho} \frac{\partial}{\partial \phi}.$$

Now, $\hat{L}_z = xp_y - yp_x$ gives us

$$\hat{L}_z = -i\hbar \left\{ \rho \cos(\phi) \left[sin\phi \frac{\partial}{\partial \rho} + \frac{\cos\phi}{\rho} \frac{\partial}{\partial \phi} \right] \right.$$

$$\left. -\rho \sin(\phi) \left[\cos\phi \frac{\partial}{\partial \rho} - \frac{\sin\phi}{\rho} \frac{\partial}{\partial \phi} \right] \right\}.$$

A spectacular simplification yields the final result:

$$\hat{L}_z = -i\hbar \frac{\partial}{\partial \phi}. \tag{13.7}$$

So, $-i\hbar \frac{\partial}{\partial \phi}$ has the fundamental commutation rule with ϕ. The eigenvalue equation

$$\hat{L}_z \psi_m(\phi) = m\hbar \psi_m(\phi) \tag{13.8}$$

yields the eigenfunctions

$$\psi_m(\phi) = \frac{e^{im\phi}}{\sqrt{2\pi}}. \tag{13.9}$$

Evidently, $\hbar$ is the fundamental unit of angular momentum, and the condition $\psi_{2\pi}(\phi) = \psi_m(0)$ requires that the *azimuthal quantum number* m be the integer, $m = 0, \pm 1, \pm 2, \ldots$

The result says that the angular momentum of any system must be a multiple.[3] Classically, it is obvious that by tilting the rotation axis a little bit, the component of L must change with continuity. Instead the measured values are discrete, while the

[3] This is the orbital angular momentum; the spin has another nature and will be discussed later.

probabilities of measuring those values vary with continuity, as we shall see. The energy eigenvalues of the rotor are

$$E_m = \frac{1}{2I} m^2 \hbar^2. \tag{13.10}$$

13.6 Commutators: Properties and Tricks

From the definition of the commutator of two operators $\hat{A}$ and $\hat{B}$, namely

$$[\hat{A}, \hat{B}]_- = \hat{A}\hat{B} - \hat{B}\hat{A},$$

it is obvious that

$$\begin{gathered}
[\hat{A}, \hat{B}]_- = -[\hat{B}, \hat{A}]_-; \\
[\hat{A}, \hat{B} + \hat{C}] = [\hat{A}, \hat{B}] + [\hat{A}, \hat{C}]; \\
[\hat{A}\hat{B}, \hat{C}]_- = \hat{A}[\hat{B}, \hat{C}]_- + [\hat{A}, \hat{C}]_-\hat{B}; \\
[\hat{A}, \hat{B}\hat{C}]_- = \hat{B}[\hat{A}, \hat{C}]_- + [\hat{A}, \hat{B}]_-\hat{C}.
\end{gathered} \tag{13.11}$$

The analogy with the Poisson brackets is also evident. Simple commutators are readily obtained.

The result

$$[x^n, \hat{p}] = i\hbar n x^{n-1} \tag{13.12}$$

can be obtained by acting over a $f(x)$ or else by starting with $[x^2, p]_- = 2i\hbar x$; otherwise, one can prove the result by induction, starting with $A = x$, $B = x^2$, $C = p$, $[x^3, p]_- = x[x^2, p] + [x, p]x^2 = 3i\hbar x^2$ and so on. From the definition of the angular momentum $\overrightarrow{L} = \overrightarrow{r} \wedge \overrightarrow{p}$

$$L_x = yp_z - zp_y, \quad L_y = zp_x - xp_z, \quad L_z = xp_y - yp_x,$$

and from the fundamental commutator $[p_z, z] = -i\hbar$, one finds

$$[L_x, L_y] = yp_x[p_z, z] + [z, p_z]p_y x = i\hbar L_z, \tag{13.13}$$

and cyclic results that can be summarized as follows:

$$\overrightarrow{L} \wedge \overrightarrow{L} = i\hbar \overrightarrow{L}. \tag{13.14}$$

Similarly, one finds

$$[\overrightarrow{L}^2, L_x] = [L_y^2, L_x] + [L_z^2, L_x],$$

$$[L_y^2, L_x] = L_y[L_y, L_x] + [L_y, L_x]L_y = -i\hbar(L_yL_z + L_zL_y) = -i\hbar\left[L_y, L_z\right]_+,$$

$$[L_z^2, L_x] = L_z[L_z, L_x] + [L_z, L_x]L_z = \mathrm{i}\hbar\left[L_y, L_z\right]_+.$$

Thus, it turns out that

$$[\overrightarrow{L}^2, L_x] = 0. \tag{13.15}$$

In other words, the square $\overrightarrow{L}^2$ of the angular momentum and any one of the components exist simultaneously.

Problem 13.1 For given operators $\hat{A}, \hat{B}, \hat{C}, \hat{D}, \hat{E}, \hat{F}$, compute

$$[\hat{A}\hat{B}\hat{C}, \hat{D}\hat{E}\hat{F}]_-.$$

Solution 13.1 One finds

$$\begin{aligned}[\hat{A}\hat{B}\hat{C}, \hat{D}\hat{E}\hat{F}]_- = {} & \hat{A}\hat{B}\{[\hat{C}, \hat{D}]_-\hat{E}\hat{F} + \hat{D}[\hat{C}, \hat{E}]_-\hat{F} + \hat{D}\hat{E}[\hat{C}, \hat{F}]_-\} \\ & + \hat{A}\{[\hat{B}, \hat{D}]_-\hat{E}\hat{F} + \hat{D}[\hat{B}, \hat{E}]_-\hat{F} + \hat{D}\hat{E}[\hat{B}, \hat{F}]_-\}\hat{C} \\ & + \{[\hat{A}, \hat{D}]_-\hat{E}\hat{F} + \hat{D}[\hat{A}, \hat{E}]_-\hat{F} + \hat{D}\hat{E}[\hat{A}, \hat{F}]_-\}\hat{B}\hat{C}.\end{aligned}$$

13.7 Angular Momentum in 3 Dimensions

13.7.1 The Algebra

From Eqs. (13.7, 13.8 and 13.9) we know that each component of $\overrightarrow{L}$ has eigenvalues $m\hbar$, with m an integer. We cannot obtain the eigenvalues of L^2 directly by summing the squares of the components, because they fail to commute. However, one can find everything about angular momentum from Eqs. (13.14 and 13.15). Together with L^2, we can have one diagonal component and we choose L_z; so there is a subspace of functions $|\lambda, m\rangle$, where λ is the (yet unknown) eigenvalue of L^2 and $L_z|\lambda, m\rangle = m\hbar|\lambda, m\rangle$. Now, $\overrightarrow{L}^2 - L_z^2 = L_x^2 + L_x^2$ is a positive quantity; hence, for a fixed eigenvalue of $\overrightarrow{L}^2$, the eigenvalues of L_z^2 must be bounded. Let us call L the largest value that m can take. Now it is evident that in the subspace with given $\overrightarrow{L}^2$, the integer m runs from $-L$ to L. But still, we have to find the eigenvalues of $\overrightarrow{L}^2$.

It is useful to introduce the so-called *shift operators* $L_\pm$, such that

$$L_+ = L_x + iL_y, \quad L_- = L_x - iL_y.$$

Since

$$[L_z, L_\pm] = \pm\hbar L_\pm,$$

one finds

$$L_z L_\pm |\lambda, m\rangle = (m \pm 1)\hbar L_\pm |\lambda, m\rangle,$$

therefore

$$L_\pm |\lambda, m\rangle = C_\pm |\lambda, m \pm 1\rangle. \tag{13.16}$$

Since λ is the maximum eigenvalue, we run into a contradiction, unless

$$L_+ |\lambda, l\rangle = 0.$$

The identity

$$L_- L_+ + L_z^2 + \hbar L_z = L^2 \tag{13.17}$$

is readily verified, and applied to $|\lambda, L\rangle$ yielding the eigenvalue of $\overrightarrow{L}^2$, namely $\lambda = L(L+1)$. However, the usage is to denote the eigenstates by $|L, m\rangle$ instead of writing $|L(L+1), m\rangle$.

Scalar multiplication of (13.16) by itself gives us

$$C_\pm^2 = [L(L+1) - m(m \pm 1)]\hbar^2,$$

and we obtain the useful result

$$L_\pm |L, m\rangle = \hbar\sqrt{L(L+1) - m(m \pm 1)}\,|L, m \pm 1\rangle. \tag{13.18}$$

13.7.2 Angular Momentum Matrices

We are able to find the matrices of $\overrightarrow{L}$ on the $|l, m\rangle$ basis. They are

$$\langle l_1 m_1 | L^2 | l_2 m_2 \rangle = \hbar^2 l_1(l_1+1)\delta_{l_1,l_2}\delta_{m_1,m_2} \tag{13.19}$$

$$\langle l_1 m_1 | L_z | l_2 m_2 \rangle = \hbar m_1 \delta_{l_1,l_2}\delta_{m_1,m_2} \tag{13.20}$$

$$\langle l_1 m_1 | L_\pm | l_2 m_2 \rangle = \hbar\sqrt{l_2(l_2+1) - m_2(m_2 \pm 1)}\,\delta_{l_1,l_2}\delta_{m_1,m_2\pm 1}. \tag{13.21}$$

Example 13.1 Matrices of di $l = 1$ in units of $\hbar$

$$L_z = \mathrm{diag}(1, 0, -1);\ L_+|1,-1\rangle = \sqrt{2}|1,0\rangle;\ L_+|1,0\rangle = \sqrt{2}|1,1\rangle;\ L_+|1,1\rangle = 0.$$

We assign the basis vectors with

$$|1,1\rangle \to \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, |1,0\rangle \to \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, |1,-1\rangle \to \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

and find (denoting the matrices with the names of the operators):

$$L_+ = \begin{pmatrix} 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{2} \\ 0 & 0 & 0 \end{pmatrix}; \tag{13.22}$$

it follows that

$$L_- = \begin{pmatrix} 0 & 0 & 0 \\ \sqrt{2} & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{pmatrix}, \tag{13.23}$$

and so

$$L_x = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \tag{13.24}$$

$$L_y = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}. \tag{13.25}$$

In general, in terms of matrices of size $2l+1$, one can represent the angular momentum over the $|l,m\rangle$ basis. The matrices of $\overrightarrow{L}$ obey the same commutation rules and the same eigenvalues as the operators; they form a *representation* of angular momentum.

13.8 Unitary Operators, Translations and Rotations

Frame changes like translations and rotations (and more) are naturally introduced as changes of the coordinates on which wave functions depend. However, in order to be able to apply Group theory and other powerful mathematical methods, it is important to be able to represent frame changes as operators acting not on the coordinates but directly on the wave function itself. We must be able to describe all phenomena from all possible viewpoints, and any kind of reference change is a quantum operator, let's say $\hat{A}$, acting on the wave function ψ.

It is easy to see that $\hat{A}$ must be *unitary*, that is,

$$\hat{A}^\dagger = \hat{A}^{-1}. \tag{13.26}$$

This property ensures that in the change of reference any normalized wave function in sent to a normalized wave function, since

$$\langle A\psi|A\psi\rangle = \langle\psi|A^\dagger A\psi\rangle = \langle\psi|\psi\rangle. \tag{13.27}$$

13.8.1 Translations

In one dimension, a translation by a is done by an operator $\hat{T}_a$ such that for any analytic function f,

$$\hat{T}_a f(x) = f(x+a). \tag{13.28}$$

Note carefully that if $a > 0$, $x \to x + a$ translates the point to the right, but the function is translated to the left. Since

$$f(x+a) = \sum_{n=0}^{\inf} \frac{a^n}{n!}\left(\frac{d}{dx}\right)^n f(x),$$

we may write

$$\hat{T}_a = \sum_{n=0}^{\inf} \frac{a^n}{n!}\left(\frac{d}{dx}\right)^n = e^{a\frac{d}{dx}} = e^{ia\frac{p}{\hbar}}. \tag{13.29}$$

This is a *hyperdifferential operator* (it does derivatives of any order up to ∞). It extends to three dimensions in the obvious way.

The momentum is the *generator* of the infinitesimal translations. It commutes with H and is conserved if the system is invariant under translation. Besides, since p is Hermitian, $\hat{T}_a^\dagger = \hat{T}_{-a}$; this verifies the unitarity.

13.8.2 Rotations

The matrix which rotates a point around the z axis by an angle ϕ is

$$M_\phi^{(z)} = \begin{pmatrix} \cos(\phi) & \sin(\phi) & 0 \\ -\sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{13.30}$$

For a point on the y axis, $(0, 1, 0) \to (\sin(\phi), \cos(\phi), 0)$; one moment's reflection shows that for $\phi > 0$, $\overrightarrow{x}' = M_\phi^{(z)}\overrightarrow{x}$ is rotated clockwise for positive ϕ; instead, the rotation of a function is defined as $f(x, y, z) \to R_\phi^{(z)} f(x, y, z)$, where

$$R_{\phi}^{(z)} f(x, y, z) = f(x\cos(\phi) + y\sin(\phi), y\cos(\phi) - x\sin(\phi), z), \tag{13.31}$$

and the same angle that rotates the point clockwise rotates R_{ϕ} counterclockwise.

The operator $R_{\phi}^{(z)}$ can be written in terms of the angular momentum operator $\overrightarrow{L}$. To see how, we consider ϕ as the effect of a succession of $n \to \infty$ infinitesimal rotations around z, $\delta\phi = \frac{\phi}{n}$. An infinitesimal rotation with $\cos(\phi) \sim 1$, $\sin(\phi) \sim \phi$, produces a change that is related to $\overrightarrow{L}_z$, since

$$f(x, y, z) \overrightarrow{R_{\delta\phi}^{(z)}} f(x', y', z) \sim f(x + y\delta\phi, y - x\delta\phi, z)$$

$$\sim f(x, y, z) + \delta\phi \left\{ y\frac{\partial f}{\partial x} - x\frac{\partial f}{\partial y} \right\} = f(x, y, z) + \delta\phi(-i)L_z f(x, y, z)$$

and

$$f(x, y, z) \overrightarrow{R_{\delta\phi}^{(z)}} \left[1 - i\delta\phi L_z\right] f(x, y, z).$$

To sum up, the operator of the infinitesimal rotation around the z axis is

$$R_{\delta\phi}^{(z)} = \left[1 - i\delta\phi L_z\right] \tag{13.32}$$

and $\boldsymbol{L}$ is the generator. For a finite rotation, we need the product

$$R_{\Delta\phi}^{(z)} = \Pi_{\delta\phi} R_{\delta\phi}. \tag{13.33}$$

By the known identity

$$e^{-a} = \underset{n \to \infty}{Lim} \left(1 - \frac{a}{n}\right)^n,$$

one obtains the operator that does a finite rotation $\Delta\phi$ around z in the form

$$R_{\phi}^{(z)} = e^{-i\frac{\Delta\phi L_z}{\hbar}}. \tag{13.34}$$

For a general rotation, one must take into account that rotations around different axes do not commute. For a rotation by an angle α around an arbitrary axis, one can write

$$R_{\alpha} = R_{\phi}^{(z)} R_{\theta}^{(x)} R_{\psi}^{(z)}, \tag{13.35}$$

where $\phi \in (0, 2\pi)$, $\theta \in (0, \pi)$ and $\psi \in (0, 2\pi)$ are a Euler angle representation of the rotation.

The set of all the operators (18.29) constitutes a mathematical structure called the O(3) Group. This means that multiplying the operators we get the operator that

corresponds to doing the rotations one after another, with each operator having an inverse. The symbol O(3) means that the rotations are in 3 dimensions and can be represented by orthogonal matrices. The use of Group theoretical methods is extremely useful in advanced Quantum Mechanics, for the classification of electron states in molecular and solid state theory, the classifications of elementary particles and many other advanced problems. Any rotationally invariant system has the property that $[H, \overrightarrow{L}]_- = 0$. The angular momentum is conserved in such cases, as in Classical Mechanics (replacing the Poisson Brackets by the commutator).

Galileo transformations are deferred to Chap. 15.

13.9 General Uncertainty Principle and Compatible Observables

Consider the measurement of some physical quantity A in a given system, like a molecule. Classically one could obtain a set of outcomes with a mean value $\langle A \rangle$, which is the expected value, and some standard deviation σ_A, which informs us about the uncertainty of the mean value. In Quantum Mechanics, the system is described by the wave function ψ, and the outcome of the experiment can be summarized by a mean value:

$$\langle \hat{A} \rangle \equiv \langle \psi | \hat{A} \psi \rangle \tag{13.36}$$

and by a standard deviation $\sigma_A = \sqrt{\langle [A - \langle A \rangle]^2 \rangle}$. A novelty of Quantum Mechanics is that the existence of a wave function which contains all the information creates an interdependence among the measurements of different observables. The precision with which two observables $\hat{A}$ and $\hat{B}$ can be measured simultaneously is limited by the following inequality[4]:

$$\sigma_A^2 \sigma_B^2 \geq \left(\frac{\langle \psi | [\hat{A}, \hat{B}]_- | \psi \rangle}{2i} \right)^2 \quad \forall \psi, \tag{13.37}$$

where $\sigma_A^2 = \langle [A - \langle A \rangle]^2 \rangle$. This implies that $\hat{A}$ and $\hat{B}$ are compatible, that is, can be both measured with precision, only if they commute. For example, momentum and position of a particle are not compatible. If ψ is well approximated by a plane wave, the uncertainty on momentum is small, but the uncertainty of position is very large. Heisenberg proposed this as a fundamental principle of Physics,[5] while Einstein disagreed strongly, and thought that a complete theory should disclose the missing information. The result shows that, in Quantum Mechanics, two observables are compatible (= measurable simultaneously and exactly) only if their operators commute.

[4] If one picks $\psi =$ eigenstate of $\hat{A}$, say, then $\sigma_A = 0$; in this limiting case σ_B blows up.

[5] In the present formulation, it is a consequence of the Schwarz inequality, as we shall see below.

To prove this, one can assume[6] that $\langle\hat{A}\rangle = 0$, $\langle\hat{B}\rangle = 0$; now, we may write $\sigma_A^2 = \langle\psi|\hat{A}^2|\psi\rangle = \langle\hat{A}^\dagger\psi|\hat{A}\psi\rangle = |\langle f|f\rangle$, where $|f\rangle = \hat{A}|\psi\rangle$, and similarly, $\sigma_B^2 = |\langle g|g\rangle$, where $|g\rangle = \hat{B}\psi\rangle$.

According to the Schwarz inequality,

$$\sigma_A^2\sigma_B^2 = \langle\psi|\hat{A}^2|\psi\rangle\langle\psi|\hat{B}^2|\psi\rangle = \langle f|f\rangle\langle g|g\rangle \geq |\langle f|g\rangle|^2 = |\langle\psi|\hat{A}\hat{B}|\psi\rangle|^2,$$

since $\langle f|g\rangle = \langle AB\rangle$. To reach the final form, we need another trivial inequality. For every complex number z,

$$|z|^2 = (Re(z))^2 + (Im(z))^2 \geq (Im(z))^2 = \left(\frac{z - z^*}{2i}\right)^2.$$

Letting $z = \langle f|g\rangle$, one obtains

$$\sigma_A^2\sigma_B^2 \geq \left(\frac{\langle f|g\rangle - \langle g|f\rangle}{2i}\right)^2.$$

Now, since $\langle f|g\rangle = \langle AB\rangle$, and $\langle g|f\rangle = \langle BA\rangle$, (13.37) results. The energy–time uncertainty deserves a separate discussion (see Sect. 15.1.1).

Minimal Uncertainty

There are no limits to ignorance: one can easily conceive states with large uncertainty both in x and p. Then, $\sigma_A\sigma_B$ is much larger than the theoretical limit. It is interesting to find out the condition on ψ that grants a minimum value of $\sigma_A\sigma_B$; it is assumed that ψ is not an eigenstate of either observable. This condition emerges if we reverse-engineer the proof in the last paragraph. The main step there was the use of Schwartz inequality on the vectors $|f\rangle = (\hat{A} - < A >)|\psi\rangle$, $|g\rangle = (\hat{B} - < B >)|\psi\rangle$; the inequality becomes an equality for parallel vectors, therefore we take $|f\rangle$ proportional to $|g\rangle$. Moreover, we inserted a $\geq$ sign when we neglected the real part of $z^2 = \langle f|g\rangle^2$, but this becomes an $=$ sign if $\langle f|g\rangle$ is imaginary. Thus, picking

$$|g\rangle = i\lambda|f\rangle, \quad \text{with real } \lambda,$$

we obtain

$$\sigma_A^2\sigma_B^2 = \left(\frac{\langle\psi|[\hat{A}, \hat{B}]_-|\psi\rangle}{2i}\right)^2.$$

This depends on the parameter λ because there is a family of states ranging from sharp B with ill-defined A to the opposite case.

[6] Indeed, $\langle\hat{A} - \langle\hat{A}\rangle\rangle = 0$ and $\sigma^2_{\hat{A}-\langle\hat{A}\rangle} = \sigma^2_{\hat{A}}$; moreover, the commutator of $\hat{A}$ and $\hat{B}$ is not changed if we subtract the mean values from the operators.

In the special case $A = \hat{p},\ B = \hat{x}$, the condition for minimum uncertainty reads

$$(\hat{p} - \langle p \rangle)\Psi = i\lambda(x - \langle x \rangle)\Psi; \tag{13.38}$$

to find out the shape of this wave function, we may shift the origin and write

$$(\hat{p} - \langle p \rangle)\Psi = i\lambda x \Psi.$$

One can see by inspection that there is a family of solutions, $\Psi = \exp[2\pi \frac{i<p>x}{h}]\Psi_0$ where

$$-i\hbar \frac{d}{dx}\Psi_0(x) = i\lambda x \Psi_0.$$

Equation (13.38) is solved by the Gaussian shifted packet (to be normalized)

$$\Psi(x) = \exp\left[-\lambda \frac{(x - < x >)^2}{2\hbar} + \frac{i < p > x}{\hbar}\right].$$

The ground state of the Harmonic oscillator is of this form, as we shall see.

Theorem *If $\hat{A}$ and $\hat{B}$ have a complete common set of eigenfunctions, they commute. Suppose the two operators $\hat{A}$ and $\hat{B}$ have a complete common set ψ_{ab} of eigenfunctions. Then,*

$$\hat{A}\psi_{ab} = a\psi_{ab}, \quad \hat{B}\psi_{ab} = b\psi_{ab} \quad \forall \psi_{ab} \implies \hat{A}\hat{B}\psi = ab\psi = \hat{B}\hat{A}\psi.$$

Any relation that holds for the full complete set is an operator relation. So,

$$\implies [\hat{A}, \hat{B}]_- = 0.$$

The matrices taken on the set ψ_{ab} are diagonal, hence the matrices also commute.

We know that a unitary operator U is such that

$$UU^\dagger = 1 = U^\dagger U. \tag{13.39}$$

Now consider the two Hermitian matrices A and B, which represent the operators on the basis ψ_{ab}. We can transform to any other basis by a unitary transformation U by writing $A \to \tilde{A} = UAU^\dagger$. Since $UAU^\dagger UBU^\dagger = UBU^\dagger UAU^\dagger \Leftrightarrow [\tilde{A}, \tilde{B}] = 0$, the two matrices commute over any basis. Then, the commuting observables are compatible, while non-commuting ones are not.

13.9.1 Canonical Transformations in Quantum Mechanics

An immediate consequence of the above comments is that a unitary transformation U preserves the commutation rules; this is like the classical canonical transformations that preserve the Poisson brackets. U gives an equivalent description of the same physics. It can be a mere change of reference, but in several cases, it is a sort of *deus ex machina* which leads to the solution of otherwise intractable problems.

Chapter 14
Postulate 3

Any measurement of an observable $\hat{Q}$ must yield an eigenvalue λ of $\hat{Q}$ (solution to the eigenvalue equation $\hat{Q}|\Psi_\lambda\rangle = \lambda|\Psi_\lambda\rangle$). The measurement prepares the system in a state in which $\hat{Q} = \lambda$. If the system is in a state $|\Phi\rangle$, the probability of a particular λ is $P(\lambda) = \langle\Phi|\Psi_\lambda\rangle|^2$. This is known as the Born Rule.

Note that the system might have a large number of degrees of freedom, yet one can make a measurement involving one of them, like one component of angular momentum, which has an eigenvalue equation depending on a single angle Φ. If $|\Phi\rangle$ is an eigenstate belonging to a discrete eigenvalue λ, $P(\lambda) = 1$; in such cases, Quantum Mechanics gives certainties. Otherwise, after the measurement Φ *collapses* in an eigenstate of the operator, and if the measurement is repeated immediately (i.e., before the system evolves) the result is again λ.

14.1 Reasons Why the Set Must Be Complete

Since the measurement must necessarily return some value, this postulate requires

$$\sum_\lambda |\langle\Phi|\Psi_\lambda\rangle|^2 = \sum_\lambda P(\lambda) = 1;$$

this demands the expansion

$$|\Phi\rangle = \sum_\lambda |\Psi_\lambda\rangle\langle\Psi_\lambda|\Phi\rangle.$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_14

Therefore, we must assume the completeness

$$\sum_{\lambda} |\Psi_\lambda\rangle\langle\Psi_\lambda| = 1.$$

The formalism works smoothly in most cases, but sometimes one must use it with a proper choice of the function space and physical vision. For example, x is an observable. We must distinguish the variable x, the operator $\hat{x}$ and the eigenvalue that is the result of a measurement; for clarity, let us call x' the eigenvalue. Letting $\varphi^{(x')}(x)$ denote the eigenfunction, the eigenvalue equation reads as

$$\hat{x}\varphi^{(x')}(x) = x'\varphi^{(x')}(x).$$

Now, $\varphi^{(x')}(x) = \delta(x - x')$ does not belong to L^2 but to the space of the distributions. This also provides the required complete set, since

$$\forall\psi, \langle\varphi^{(x)}|\psi\rangle = \int dx'\varphi^{(x)}(x')\psi(x') = \psi(x).$$

But $\delta(x - x')$ is not a normal wave function, and, for instance, we cannot take the square modulus. The difficulty stems from the fact that x is a continuous variable, and it is impossible to make a measurement giving the position of a particle quite exactly; that would require infinite energy.

With this proviso, we must be ready to work with discrete and continuous observables. In the discrete case, the eigenvalue equation reads as $\hat{Q}|e_n\rangle = \lambda_n|e_n\rangle$, with $n = 1, 2, 3, \ldots$, and $\langle e_m|e_n\rangle = \delta_{mn}$; besides, $|\Psi\rangle = \sum_n c_n|e_n\rangle$ with $c_n = \langle e_n|\Psi\rangle$. Then, $P(n) = |c_n|^2$ is the probability of the eigenvalue λ_n. One notable example is provided by angular momentum. Any direction can be taken as the z axis, and $\hat{L}_z$ is the orbital angular momentum operator.[1] Any measurement must return $\frac{hm}{2\pi}$, where the azimuthal quantum number m is an integer.

In the continuous case, the eigenvalue equation reads as $\hat{Q}|e_k\rangle = \lambda_k|e_k\rangle$, with $-\infty < k < \infty$, but $\langle e_h|e_k\rangle = \delta(h - k)$; besides, $|\Psi\rangle = \int dk c_k|e_k\rangle$, where $c_k = \langle e_k|\Psi\rangle$. So, $dP = |c_k|^2 dk$ is the probability of finding the eigenvalue within dk from λ_k.

We have just discussed the operator $\hat{x}$; as another example, we may take the momentum $\hat{p}$, having eigenfunctions $e_p(x) = \frac{e^{ikx}}{\sqrt{2\pi}}$, where $p = \hbar k$. The p component of $|\Psi\rangle$ is $c_p = \langle e_p|\Psi\rangle = \int_{-\infty}^{\infty} e_p(x)\Psi(x)dx = \Psi(p, t)$; this is the wave function in momentum space.

[1] The spin angular momentum will be introduced in Chap. 18.

Chapter 15
Postulate 4

The time evolution of ψ is governed by the Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = \hat{H}(t)\psi(t), \tag{15.1}$$

where $\hat{H}(t)$ is the Hamiltonian.

15.1 Time Derivative of an Observable and F = ma

I anticipate here that this statement (like all the postulates) holds for single-particle and many-particle systems as well (although in the latter case we have to specify in later chapters how the equation works in detail).

The Schrödinger equation is first-order in time; therefore, if ψ is given at one time, it can be evolved in the future or in the past at all times. A second initial condition that is necessary in the classical case is not needed. It is also second-order in space variables, which implies that the theory is not acceptable from the standpoint of Relativity. Proper extensions (Klein–Gordon equation for spinless particles, Dirac equation for spin $\frac{1}{2}$ Fermions, Maxwell equations for Photons, the Proca equation for massive spin-1 particles and so on) are known since a long time; they are quite important, but do not spoil the relevance of the non-relativistic approximation to many interesting phenomena.

Note that even the matrix elements of time-independent operators generally depend on the time because of the evolution of the wave function. The derivative $\frac{d\hat{Q}}{dt}$ of an operator $\hat{Q}$ is by definition an operator that satisfies

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_15

$$\left\langle \Psi | \frac{d\hat{Q}}{dt} | \Psi \right\rangle \equiv \frac{d}{dt} \left\langle \Psi | \hat{Q} | \Psi \right\rangle = \left\langle \Psi | \frac{\partial}{\partial t} \hat{Q} | \Psi \right\rangle + \left\langle \frac{\partial}{\partial t} \Psi | \hat{Q} | \Psi \right\rangle + \left\langle \Psi | \hat{Q} | \frac{\partial}{\partial t} \Psi \right\rangle.$$

Using the Schrödinger equation one finds $\hbar|\dot{\psi}\rangle = -iH|\psi\rangle \Longrightarrow \hbar\langle\dot{\psi}| = i\langle\psi|H$, and so,

$$\frac{d\hat{Q}}{dt} = \frac{\partial \hat{Q}}{\partial t} + \frac{i}{\hbar}[\hat{H}, \hat{Q}]. \tag{15.2}$$

We know that in Classical Physics, the evolution of a function F of the canonical variables is given by a similar equation, namely $\dot{F} = \frac{\partial F}{\partial t} + \{F, H\}$. The commutators replace the Poisson brackets and, if $\frac{\partial \hat{Q}}{\partial t} = 0$, the operators that commute with the Hamiltonian are conserved. For example, consider a particle in a potential; using bold face for the vectors, $\hat{H} = \frac{\hat{\mathbf{p}}^2}{2m} + V(x)$. The velocity is $\frac{d\hat{\mathbf{x}}}{dt} = \frac{i}{\hbar}[\hat{H}, \hat{\mathbf{x}}] = \frac{\mathbf{p}}{m}$, as in Classical Physics. Thus, $m\frac{d^2\hat{x}}{dt^2} = \frac{i}{\hbar}[\hat{H}, \frac{\hat{p}}{m}] = [V, \nabla] = \nabla V$. So, we find

$$m\ddot{\mathbf{x}} = \hat{\mathbf{F}}, \tag{15.3}$$

where $\hat{\mathbf{F}} = -\nabla V$ is the force. Wow! We have derived the Newtonian equation of motion. It is still valid! Albeit, with a different interpretation, of course. This is a recurrent phenomenon: we shall see that several important classical equations (including Maxwell's equations) remain valid in Quantum Theory even if they must be interpreted in a new way.

15.1.1 Energy–Time Uncertainty

The uncertainty principle of Sect. 13.9 does not involve the time, since there is no Hermitian operator for t. However, consider the uncertainty principle,

$$(\Delta A\ \Delta B)^2 \equiv \sigma_A^2 \sigma_B^2 \geq \left(\frac{\langle\psi|[\hat{A}, \hat{B}]|\psi\rangle}{2i} \right)^2,$$

for $A = \hat{H}$, the Hamiltonian, $B = \hat{Q}$, where $\hat{Q}$ is a time-independent observable. The commutator with H appears in (15.2), so one finds that

$$\Delta E \Delta Q \geq \frac{1}{2}\hbar| < \Psi | \frac{dQ}{dt} | \Psi > |. \tag{15.4}$$

The derivative of Q is due to the evolution of the wave packet. If for $|\Psi >$, we choose an eigenstate of H, there is no evolution, and the derivative vanishes. For a system with a continuous spectrum, this never happens; experimentally, $|\Psi >$ is always a wave packet. Suppose that someone tries to devise an experiment in which the wave

packet is prepared and then Q is measured so fast that the evolution barely begins. Then,

$$\Delta Q \approx < \frac{dQ}{dt} > \Delta t.$$

Putting this into (15.4), one finds

$$\Delta E \Delta t > \frac{1}{2}\hbar. \tag{15.5}$$

One cannot have a small Δt and simultaneously an arbitrarily small ΔE; being fast requires mixing different energies. Conversely, a well-defined energy requires an experiment of sufficient length. This is the energy–time uncertainty principle. Mathematically, this principle reflects known properties of the Fourier transform. A finite piece of a sinusoid mixes all frequencies.

Only long-lived atomic states have relatively sharp energies; therefore, atomic clocks use so-called *forbidden* transitions, which actually take place but take long time. Several years ago, an error of one second in one million years was claimed. In 2015 a metrology laboratory at Riken in Japan has built a clock which misses a second in 15 billion years, which is longer than the age of the Universe.

Problem 15.1 A particle is in the ground state of an oscillator with Hamiltonian $H[1] = \frac{p^2}{2m} + \frac{1}{2}m\omega_0^2 x^2$. Calculate the probability $P(\eta)$ that it is in the ground state of a harmonic oscillator with Hamiltonian $H[\eta] = \frac{p^2}{2m} + \frac{1}{2}m(\eta\omega_0)^2 x^2$.

Solution 15.1 Using

$$\psi_0(x) = \frac{1}{\sqrt{x_1\sqrt{\pi}}} e^{-\frac{x^2}{2x_1^2}},$$

with $x_\eta^2 = \frac{\hbar}{m\eta\omega} = \frac{x_1}{\eta}$, letting $\frac{1}{x_a^2} = \frac{1}{2}(\frac{1}{x_1^2} + \frac{1}{x_\eta^2})$, one finds

$$\langle\psi_\eta|\psi_1\rangle = \frac{\sqrt{\pi}x_a}{\sqrt{\pi x_1 x_\eta}} = \sqrt{2}\frac{\eta^{\frac{1}{4}}}{\sqrt{1+\eta}}$$

and therefore

$$P(\eta) = 2\frac{\sqrt{\eta}}{1+\eta} = \frac{2}{\sqrt{\eta} + \frac{1}{\sqrt{\eta}}}.$$

15.1.2 Adiabatic and Quasi-Adiabatic Evolution

A quantum system in a stationary state of a constant Hamiltonian H has no expiration: once it occupies an eigenstate it remains there forever. A time-dependent Hamiltonian $H(t)$ cannot have stationary states and is much more difficult to deal with. However,

intuition suggests that if the time dependence is slow enough, in a sense that needs to be precisely defined, one can think about transitions between slowly changing almost-stationary states. This idea has been very far-reaching. The adiabatic theorem (Kato 1949) deals with the solution of the Schrödinger equation with a time-dependent Hamiltonian $H(t)$ in the case of a discrete and non-degenerate spectrum. The theorem says that if the system is prepared in the nth eigenstate and the evolution of H is sufficiently slow, then the system will remain forever in the time-dependent nth eigenstate. The Kato theorem has many important and far-reaching consequences. The reader could be surprised by this statement, since the requirement that the spectrum be discrete and non-degenerate looks very restrictive. In fact, this limitation can be circumvented by a method introduced by Gell-Mann and Low, and the many-body Green's function theory of the electron liquid has been formulated as if the Coulomb interaction were turned on adiabatically, starting from a non-interacting system in the far past. Other reasons why adiabatic treatments are important will be apparent in Chap. 23. Since the theorem is so plausible I omit the proof, which may be found elsewhere.[1]

It is important to investigate the relation between the solution of the time-dependent Schrödinger equation and the complete set of solutions of the **stationary state** equation with the *instantaneous Hamiltonian* $H(t)$, namely

$$H(t)\phi_n(t) = \epsilon_n(t)\phi_n(t). \tag{15.6}$$

Let us consider the evolution between time 0 and time t. Substituting the formal expansion

$$\psi(t) = \sum_n \alpha_n(t)\phi_n(t) \exp\left[\frac{-i}{\hbar}\int_0^t \epsilon_n(t')dt'\right], \tag{15.7}$$

we readily find the condition

$$\sum_n (\dot{\alpha}_n\phi_n + \alpha_n\dot{\phi}_n) \exp\left[\frac{-i}{\hbar}\int_0^t \epsilon_n(t')dt'\right] = 0. \tag{15.8}$$

To find the coefficients, we take the scalar product with the instantaneous $\phi_k(t)$. We obtain

$$\dot{\alpha}_k = -\sum_n \alpha_n \langle\phi_k|\dot{\phi}_n\rangle \exp\left[\frac{-i}{\hbar}\int_0^t (\epsilon_n(t') - \epsilon_k(t'))dt'\right]. \tag{15.9}$$

We need $\langle\phi_k|\dot{\phi}_n\rangle$; to this end, we differentiate (15.6) and scalar-multiply by $\phi_k(t)$, for $k \neq n$. The result is

$$\langle\phi_k|\dot{\phi}_n\rangle = \frac{\langle\phi_k|\frac{\partial H}{\partial t}|\phi_n\rangle}{\epsilon_n - \epsilon_k}. \tag{15.10}$$

[1] The original paper by Kato is somewhat involved, but for a simple proof see D. J. Griffiths, "Introduction to Quantum Mechanics", Prentice Hall.

This result allows us to estimate the probability amplitude to jump to different states when the adiabatic condition is weakly violated. We see that if $\langle\phi_k|\frac{\partial H}{\partial t}|\phi_n\rangle$ is small compared to the energy separation $\epsilon_n - \epsilon_k$, the system has little chance to be found in k if it is prepared originally in n. Now it is verified that if the spectrum is discrete and non-degenerate, a finite gap must be overcome before the system can make any transition; therefore, any evolution can be done so slowly that the system remains in n all the time.

15.1.3 Sudden Approximation

Suppose one knows the (one-body or many-body) normalized wave function $\psi(t_0)$ of some system and the Hamiltonian $H(t_0)$ at some time t_0; if H stays constant for $t > t_0$, we can write down immediately the evolution of $\psi(t_0)$ in terms of the eigenfunctions ϕ_n of H at time t_0, setting $\hbar = 1$,

$$\psi(t) = \sum_n \phi_n e^{-i\epsilon_n(t-t_0)}\langle\phi_n|\psi\rangle, \tag{15.11}$$

where the sum can imply an integration in the case of a continuous spectrum. Typically, this result is useful if $\psi(t_0)$ is the wave function resulting from the evolution of the system for times before t_0 with a different Hamiltonian H'. This means that we are considering a model in which the Hamiltonian changes abruptly at time t_0, inevitably leaving the system in a mixture of eigenstates of the new Hamiltonian. While it is clear that Hamiltonians cannot really change instantly, there are many situations in which such a scheme is very useful. For instance, when a many-electron atom in its ground state suffers a nuclear decay that leads to a change of the atomic number, the change is fast enough to be considered as practically instantaneous when one wishes to calculate the probability that the atom gets ionized in the process; indeed, the characteristic electronic times are long compared to those of the much more energetic nuclear transitions. In addition, the sudden approximation offers a practical scheme for the solution of arbitrary time-dependent problems by an iteration of the above idea. Instead of solving the problem with $H(t)$ for $t \in (t_i, t_f)$, one divides the time interval, introducing intermediate times such that $t_i < t_1 < t_2 < \cdots < t_f$, approximating H(t) with a constant Hamiltonian in each sub-interval. If each sub-interval is short compared to the characteristic times of the system and the change of the Hamiltonian, this is a convenient, practical way to compute the result.

Problem 15.2 A particle of mass m in 1 dimension is in the ground state of the potential well:

$$V(x) = \begin{cases} \infty \text{ per } x < -\frac{a}{2}, \\ 0 \quad \text{per } 0 < x < a, \\ \infty \text{ per } x > \frac{a}{2}. \end{cases}$$

At time t = 0, the potential becomes $V(x) = 0$ everywhere. Write the wave function $\psi(x, t)$ at all times $t > 0$.

Solution 15.2 Since $\psi_0(x) = \sqrt{\frac{2}{a}} \sin(\frac{\pi(x-\frac{a}{2})}{a})$ and the plane wave is

$$\phi_k(x, t) = \frac{e^{i(kx-\omega(k)t)}}{\sqrt{2\pi}},$$

with

$$\hbar\omega = \frac{(\hbar k)^2}{2\,m},$$

the expansion $\psi_0(x) = \int dk \phi_k(x) \langle \phi_k | \psi_0 \rangle$ requires that

$$\langle \phi_k(x, 0) | \psi_0(x) \rangle = \alpha(k).$$

Using

$$\alpha(k) = \int_{-\frac{a}{2}}^{\frac{a}{2}} dx \frac{e^{-ikx}}{\sqrt{\pi a}} \sin\left(\frac{\pi(x-\frac{a}{2})}{a}\right) = 2\sqrt{\pi a} \frac{\cos(\frac{ak}{2})}{a^2k^2 - \pi^2},$$

one obtains

$$\psi(x, t) = \int_{-\infty}^{\infty} \alpha(k) \frac{e^{i(kx-\omega(k)t}}{\sqrt{2\pi}} dk = \sqrt{\frac{a}{2}} \int_{-\infty}^{\infty} \frac{\cos(\frac{ak}{2})}{a^2k^2 - \pi^2} e^{i(kx-\omega(k)t)} dk.$$

15.1.4 Connection to Statistical Mechanics

Consider a one-dimensional system, maybe a chain of sites, with a time-independent Hamiltonian $\hat{H}$ and its evolution from an initial state $\psi(0)$ in which the particle is localized at site x. The Schrödinger equation is solved by

$$\psi(t) = \exp(-\frac{iHt}{\hbar}) |\psi_0\rangle \tag{15.12}$$

and so the probability amplitude to find the particle at x at time t is

$$\langle \psi(t) | \psi(0) \rangle = \langle x | \exp(-\frac{iHt}{\hbar}) | x \rangle. \tag{15.13}$$

Summing over x gives us

$$\int dx \langle x | \exp(-\frac{iHt}{\hbar}) | x \rangle \tag{15.14}$$

the trace of the evolution operator

$$\hat{U}(t) = \exp(-\frac{i\hat{H}t}{\hbar})$$

(the integral becomes a sum for a discrete model), but this is not that interesting here. The interesting fact is that complex exponential becomes real if we decide to work with an imaginary time, and setting

$$it = \hbar\beta \equiv \frac{\hbar}{K_B T} \tag{15.15}$$

we obtain the partition function of our toy model,

$$Z = Tre^{-\beta\hat{H}}. \tag{15.16}$$

The great importance of this fact is that it is completely general. Therefore, statistical mechanics and dynamical evolution are very much the same problem. Indeed, in the Kadanoff–Baym formalism and in the Keldysh formalism one computes Green's functions at finite temperature by evolving the system along general paths.[2]

15.2 Supersymmetry in Quantum Mechanics

Supersymmetry is a conjecture formulated about 1970 and relating each elementary particle to a (never observed) partner, and is part of other conjectures about imaginary Planck-sized objects called strings.[3] In the present context, however, supersymmetry is a simple method for transforming any solved problem into another problem, thus generating new exactly solved problems in Quantum Mechanics, and I intend to present it in the simplest way.[4]

Suppose we can solve exactly the Schrödinger equation with the Hamiltonian

$$H_1 = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V_1(x); \tag{15.17}$$

[2] See, for instance, "Topics and Methods in Condensed Matter Theory", by the present author, Springer Verlag.

[3] It could be a prophecy or the most important and enduring blunder in the history of theoretical physics, and the beginning of the abandonment of the Galilean method by a subset of theorists. The fact is that its proponents are completely indifferent to the experimental data and do not predict anything observable.

[4] For this section I used, among others, the Thesis by Chiara Veronese, Universitá di Padova, a.a. 2016–2017, and the Thesis by Simone Notarnicola, Universitá di Bari, a.a. 2009–2010. In the latter case, several explicit examples are presented.

by a shift of energies the lowest eigenvalue can always be taken to be $E_0 = 0$. We factor H_1 in the form

$$H_1 = A^\dagger A \tag{15.18}$$

where A is a new operator; the new (Hermitian) Hamiltonian will be in the form

$$H_2 = A A^\dagger. \tag{15.19}$$

Now, we have two quite different models, with the eigenvalue problems

$$H_1 \psi_n^{(1)} = E_n^{(1)} \psi_n^{(1)}, \tag{15.20}$$

$$H_2 \psi_n^{(2)} = E_n^{(2)} \psi_n^{(2)}, \tag{15.21}$$

but since they are supersymmetric partners, we shall see shortly that a solution to one of the two also solves the other. We set

$$A = \sqrt{\frac{\hbar^2}{2m}} \frac{d}{dx} + W(x), \tag{15.22}$$

where $W(x)$ is the superpotential, to be found. Of course,

$$A^\dagger = -\sqrt{\frac{\hbar^2}{2m}} \frac{d}{dx} + W(x). \tag{15.23}$$

In order to find $W(x)$, one can start from Eq. (15.18)

$$0 = \langle \psi_n | H_1 \psi_0 \rangle = \langle A\psi_n | A\psi_0 \rangle. \tag{15.24}$$

In particular, for n = 0,

$$0 = \langle A\psi_0 | A\psi_0 \rangle. \tag{15.25}$$

This implies

$$A|\psi_0\rangle = 0. \tag{15.26}$$

Thus, the superpotential is

$$W(x) = -\frac{\hbar}{\sqrt{2m}} \frac{\frac{d}{dx}\psi_0(x)}{\psi_0(x)}. \tag{15.27}$$

Using (15.22) and (15.23) one derives $H_1 = A^\dagger A$ and finds the Riccati equation for the unknown $W(x)$ in terms of the known potential $V_1(x)$:

$$V_1(x) = W(x)^2 - \frac{\hbar}{\sqrt{2m}} W'(x). \tag{15.28}$$

The Riccati equation, being first-order, can be considered to be easy, and in practice a solved problem; so, we can consider the superpotential as a known function. From $H_1 = A^\dagger A$, the ground state of H_1 with null eigenvalue is found from $A\psi_1^{(0)} = 0$. This gives

$$\psi_0^{(1)} = exp(-\frac{\sqrt{2\,m}}{\hbar}\int^x w(y)dy), \tag{15.29}$$

up to a normalization factor. Now, the solution for the supersymmetric partner is imminent. From Eqs. (15.19) and (15.23) we get

$$V_2 = W^2 + \frac{\hbar}{\sqrt{2\,m}}W'(x). \tag{15.30}$$

Hence,

$$V_2 - V_1 = 2\frac{\hbar}{\sqrt{2\,m}}W'(x). \tag{15.31}$$

The problems (15.20) and (15.21) have eigenfunctions $\psi_m^{(1)}$ and $\psi_m^{(2)}$, respectively, which are related since

$$H_2 A\psi_m^{(1)} = AA^\dagger A\psi_m^{(1)} = E_m^{(1)} A\psi_m^{(1)},$$

so, $A\psi_m^{(1)}$ are eigenstates of H_2, with the exception $A\psi_0^{(1)} = 0$. In the same way, $A^\dagger\psi_m^{(2)}$ are eigenstates of H_1. Moreover, the two supersymmetric Hamiltonians have the same spectrum, except for the lowest eigenvalue which is 0 for H_1 and is missing for H_2. In other terms, $E_0^{(1)} = 0$, $E_0^{(2)} = E_1^{(1)}$, $E_1^{(2)} = E_2^{(1)}$, $E_3^{(2)} = E_4^{(1)}, \ldots$.

15.3 Galileo Transformation

In Chap. 6, we performed thought experiments using the station and the train frames to introduce the Lorentz transformation. Here, we can use the same arrangement to transform the wave function in the non-relativistic case. If $\psi_S(\boldsymbol{x}_S, t)$ is the wave function of a particle in an inertial reference (the station), how is it related to the wave function $\psi_T(\boldsymbol{x}_T, t)$ in a train moving with speed $\boldsymbol{v}$?

Let us start considering a free particle in the station with momentum $\boldsymbol{p}_S$ and energy $\epsilon(\boldsymbol{p}_S)$. The Galileo transformation $\boldsymbol{x} = \boldsymbol{x}_T + \boldsymbol{v}t$ applies. For a free particle, $\boldsymbol{p}_S = \boldsymbol{p}_T + m\boldsymbol{v}$, which implies that $\epsilon(\boldsymbol{p}_S) = \epsilon_{\boldsymbol{p}_T} + \boldsymbol{v}.\boldsymbol{p}_T + \frac{1}{2}mv^2$.

This leads to

$$\psi(\boldsymbol{x}_S, t) = \psi_T(\boldsymbol{x}_T, t) \exp\left[\frac{im}{\hbar}\boldsymbol{v}.\boldsymbol{x}_T\right] \exp\left[\frac{imv^2t}{2}\right]. \tag{15.32}$$

Thus, the transformation of the wave function requires a phase factor $\exp[\frac{imv^2t}{2}]$.

Wow! Since this is p_S-independent, the same law applies to any wave packet.

In the presence of a potential, the transformation still applies in the present form if the potential is scalar, that is, independent of the reference. Otherwise, one should calculate the wave packet with the potential in the system of the station and then transform the result to the train. We have found one more use for phase factors.

15.4 Quantum Particle in Electromagnetic Field

The Schrödinger equation for a charged particle is

$$H\psi = \left(\frac{(\boldsymbol{p} - \frac{e\boldsymbol{A}}{c})^2}{2\,m} + eV\right)\psi = i\hbar\frac{\partial}{\partial t}\psi, \tag{15.33}$$

where $\boldsymbol{A}$ and V are the potentials, familiar from Classical Mechanics, the same as in Eq. (2.101). However, important changes from the classical case are in order.[5] In the classical equation of motion Lorentz force requires the knowledge of position and velocity of the particle, while they do not exist simultaneously in Quantum Mechanics. Another important difference is that classical equations of motions depend on the fields; therefore, the gauge invariance of the theory is trivial. Instead, Eq. (15.33) does not depend on the fields; it depends on the unobservable potentials that are largely arbitrary. Moreover, the unknown of Eq. (15.33) is the complex wave function ψ, and one could reasonably expect that much of the information about the state of the particle must reside on the phase: yet, one can multiply ψ by i or any other constant phase factor, and nothing changes.

But let us see what happens if we multiply the solution ψ of Eq. (15.33) by the phase factor

$$\exp(\frac{ie\chi(\mathbf{r}, \mathbf{t})}{c\hbar})$$

where χ is an arbitrary function of space–time; in this way we can change the phase of ψ in any way. Now,

$$(-i\hbar\nabla - \frac{e\mathbf{A}}{c}) \exp(i\frac{e\chi}{\hbar c})\psi \rightarrow \exp(i\frac{e\chi}{\hbar c})(\frac{-e}{c}\nabla\chi - \frac{eA}{c} - i\hbar\nabla\psi)$$

[5] Spin will produce further changes, but here I am considering a spinless particle.

which is just the gauge change

$$\boldsymbol{A} \to \boldsymbol{A} + \nabla\chi.$$

Moreover,

$$i\hbar\frac{\partial}{\partial t}(\exp\frac{ie\chi(\mathbf{r},\mathbf{t})}{\mathbf{c}\hbar}\psi) = i\hbar\exp\frac{ie\chi(\mathbf{r},\mathbf{t})}{\mathbf{c}\hbar}(\frac{-e}{c}\frac{\partial\chi}{\partial t} + \frac{\partial\psi}{\partial t});$$

this produces the gauge change

$$eV \to eV - \frac{e}{c}\frac{\partial\chi}{\partial t}. \tag{15.34}$$

Thus, we conclude that the phase of the wave function is arbitrary, but a change of the phase corresponds to a change of gauge, which is also arbitrary.

In the classical equations of motion, only the fields appear and the gauge invariance of the theory is obvious; instead, the wave function is affected by the change; however, the matrix elements of the coordinates calculated with ψ are identical to those calculated with $\psi' = \psi\exp(\frac{ie\chi(\mathbf{r},\mathbf{t})}{\mathbf{c}\hbar})$; and those of the mechanical momentum $\boldsymbol{p} - \frac{e}{c}\boldsymbol{A}$ calculated with ψ are identical with those of the mechanical momentum $\boldsymbol{p} - \frac{e}{c}(\boldsymbol{A} + \nabla\chi)$ calculated with the transformed ψ'. Consequently, the Physics is unaffected by the change.

15.4.1 The Local Phase Invariance Generates All of Electrodynamics!

In the relativistic quantum theory, one can derive everything about the Maxwell equations by just requiring that the matter field (or the wave function) couples to the charged particles in such a way that we remain free to modify the phase of the wave function independently for each time and each position is space (U(1) invariance). This fundamental fact, via the Yang–Mills theory, led to the quantum theory of weak and strong interactions and to the electroweak unification. This is one of the most beautiful results of all Theoretical Physics. However the present book is *elementary*, so it does not contain Dirac's relativistic quantum theory which is necessary for these developments. The interested reader is adviced to study David Griffiths' Introduction to Elementary Particles, Wiley, Second Edition, paragraph 10.3. However, the quantum rules that we have discussed above already tell us much about the physical meaning of the phase of charged particles in Quantum Mechanics and its relation to the electromagnetic field. While the phase at a given point in space–time is arbitrary, the space–time dependence of the phase follows the choice of the electromagnetic potentials.

15.5 Dirac's Magnetic Monopole

In 1931 Dirac investigated the possibility that some as yet undiscovered particle had a magnetic (rather than electric) charge. This would make the electromagnetic theory more symmetric.[6] Such a speculative investigation produced interesting results. Then assume there is an elementary magnetic charge g such that a point classical particle at the origin of coordinates produces a magnetic field according to the law

$$\overrightarrow{B} = \frac{g}{r^2}\overrightarrow{e}, \tag{15.35}$$

where $\overrightarrow{e}$ is a unit vector in the direction of observation. The total magnetic flux through a spherical surface centered at the particle is $4\pi g$. The analogy with electrostatics ends here. Since $\overrightarrow{B}$ is a magnetic field, we must also be able to write a vector potential $\overrightarrow{A}$.

Using the conventional spherical coordinates with longitude and latitude given by ϕ, θ, as usual, we expect to find that the integral over the parallel ϕ_0 of a sphere having radius r is

$$\int_C \overrightarrow{A} \cdot d\overrightarrow{r} = \phi_{enclosed}, \tag{15.36}$$

where $\phi_{enclosed}$ is the flux through the cap from the North Pole to the latitude ϕ. The flux through the cap is $\frac{g}{r^2} \times$ the area of the cap, which is

$$2\pi r^2(1 - \cos(\theta)).$$

The length of the parallel is $2\pi r \sin(\theta)$. So, the vector potential is

$$\overrightarrow{A} = e_\phi \frac{g(1 - \cos(\theta)}{r \sin(\theta)}, \tag{15.37}$$

where e_ϕ is a unit vector along the azimuthal direction. However, we find unexpected trouble, because for $\theta \to \pi$ our expression must collect all the flux, but the length of the parallel vanishes. In other terms, this expression is singular at $\theta = \pi$. The best that we can do is to restrict the above expression to the safe North Hemisphere, and write

$$\overrightarrow{A}_+ = e_\phi \frac{g(1 - \cos(\theta)}{r \sin(\theta)} \tag{15.38}$$

for the north hemisphere and

[6] This is an aesthetic criterion, and philosophical criteria, although unjustified, are useful. They very often lead to results just opposite to those expected. For example, Pauli was sure that parity was an exact symmetry, but the experiment says that this is not true in weak interactions. And the same happened with time reversal.

$$\overrightarrow{A}_{-} = -e_{\phi}\frac{g(1+\cos(\theta)}{r\sin(\theta)} \tag{15.39}$$

for the South. Now assume that the flux is coming from infinity to the singularities using a flux tube (called Dirac string). Can we justify the fact that no such flux tube has been observed? A particle going around it would acquire a Bohm–Aharonov phase factor $\exp(\frac{ie\phi}{\hbar})$. Here the flux is $4\pi g$. So we must require

$$\frac{4\pi eg}{\hbar} = 2\pi m, \tag{15.40}$$

with integer m. But this is striking. It implies that

$$eg = \frac{m\hbar}{2}. \tag{15.41}$$

This is interpreted to mean that the existence of a Dirac monopole would explain why the charge of elementary particles is quantized. However no monopole has been detected so far, despite many attempts.

15.6 No Cloning Theorem

The concept of probability in quantum theory refers to a distribution of results of measurements on many samples. The need for many identical samples is evident since the act of measurement collapses the wave function. This need could be bypassed if we could clone a given sample, with wave function $|\phi\rangle$, i.e., if we could produce many copies in the same quantum state without knowing it. The copies would conceivably be obtained by starting with systems having wave functions $|\alpha_i\rangle, i = 1, \ldots N$, which depend on the same variables as $|\phi\rangle$.

A quantum cloning machine is defined to operate in analogy with a photocopy machine, except that the contents of the original page are assumed unknown. Initially one has an unknown quantum state $|\phi\rangle$. For the sake of argument I assume here that it is a spin, but it could be anything. Besides, one has n spins $|\alpha_i\rangle, i = 1, \ldots n$ which have the same role as the white sheets. Here, E is the state of the machine and the environment. So, the initial state of the system at time $t = 0$ is

$$|s_{\phi}\rangle = |E\rangle|\phi\rangle|\alpha_1\rangle \cdots |\alpha_n\rangle.$$

The machine Hamiltonian has an evolution operator U such that at the end, we must have

$$U(t)|s_{\phi}\rangle = U(t)|E\rangle|\phi\rangle|\alpha_1\rangle \cdots |\alpha_n\rangle = |E_{\phi}\rangle|\phi\rangle|\phi\rangle \cdots |\phi\rangle,$$

where E_ϕ is the final state of the machine and the environment. In 1982 Wootters and Zurek[7] showed that it is impossible to make such a machine. The essential reason is that one starts with a state α such that $\langle\phi|\alpha\rangle < 1$, and we want to evolve it until $\langle\phi(t)|\alpha(t)\rangle = 1$; however $\langle\phi(t)|\alpha(t)\rangle = \langle\phi(0)|U^\dagger(t)U(t)\alpha(0)\rangle$ where U is the evolution operator; but the evolution is unitary and so $\langle\phi(t)|\alpha(t)\rangle = \langle\phi(0)|\alpha(0)\rangle$. This simple argument is incomplete because it does not take into account the evolution of the machine and the environment. Therefore the proof is slightly longer. Suppose that besides $|\phi\rangle$ we also clone a second state $|\psi\rangle$. The initial state would be

$$|E\rangle|\psi\rangle|\alpha_1\rangle\cdots|\alpha_n\rangle$$

and the final state

$$U(t)|s_\psi\rangle = U(t)|E\rangle|\psi\rangle|\alpha_1\rangle\ldots|\alpha_n\rangle = |E_\psi\rangle|\phi\rangle|\phi\rangle\cdots|\phi\rangle.$$

The overlap would be $\langle s_\phi|s_\psi\rangle = \langle\phi|\psi\rangle$ at $t = 0$; finally it should become $\langle s_\phi|U^\dagger U|s_\psi\rangle = \langle E_\phi|E_\psi\rangle\langle\phi|\psi\rangle^n$. Since U is unitary, the overlap is the same as the initial one. Hence we get

$$\langle\phi|\psi\rangle = \langle E_\phi|E_\psi\rangle\langle\phi|\psi\rangle^n \to 1 = \langle E_\phi|E_\psi\rangle\langle\phi|\psi\rangle^{n-1}$$

for any n, and this cannot be true. Therefore, U does not exist.

15.7 Path Integral Formulation of Quantum Mechanics, Part1: Particle Without Spin

The solution of partial differential equations by a technique based on paths was initiated by Robert Wiener.[8] At the end of the 40 s, Richard Feynman[9] proposed the *path integral* formulation of Quantum Mechanics, based on a remark by Dirac. It is equivalent to the Heisenberg and Schrödinger formulations, but reveals in a clearer way the connection to Classical Mechanics. It is important in field theory too, but for simplicity, I discuss its application to one-particle problems.

We introduced the quantum evolution operator $U(0, t)$ from time 0 to time t in Eq. (10.8); now, consider the amplitude $A(t) = \langle a|U(0, t)|b\rangle$ for the particle to go from state a at time $t = 0$ to a state b at time t. By using the Group property (10.11)

[7] Wootters and Zurek (1982).

[8] Robert Wiener (Boston, Massachussets, June 26, 1894-March 18, 1950) made important contributions in various areas of Mathematics including potential theory, Fourier series and integral equations. His most important contributions are the development of the mathematical theory of the Brownian motion, stochastic calculus, harmonic analysis and ergodic theory.

[9] Richard Feynman (New York City 1918–1988) won the Nobel prize in 1965 for the invention of his diagrams and his contributions to Quantum Electrodynamics.

of the evolution operator, one can slice the time interval $(0, t)$ finely to $N \gg 1$ subintervals of small duration ϵ, and the limit $N \to \infty, \epsilon \to 0$ is understood at the end of the procedure. For small ϵ, one can safely assume that H is constant within each slice even if $H = H(t)$. Consequently, by the reasoning that led us to Eq. (10.12), $U(n\epsilon, (n+1)\epsilon) = \exp(-iH\epsilon)$, where H is understood to be taken at the time slice.

The result is a terrific multiple integral:

$$A(t) =$$

$$\int dq_1 \ldots \int dq_{n-1} \langle a|e^{-iH\epsilon}|q_{n-1}\rangle\langle q_{n-1}|e^{-iH\epsilon}|q_{n-2}\rangle \cdots \langle q_2|e^{-iH\epsilon}|q_1\rangle\langle q_1|e^{-iH\epsilon}|b\rangle.$$

Here, the intermediate states q_i are taken to be position eigenstates that are a complete set, and in the limit $\epsilon \to 0$ this expression is a path integral. $A(t)$ is a sum of all the contributions that correspond to possible choices of the values of q_n at the nth slice, so it is a sum over *stories*. For each story, we can also assign the quantum particle a formal instantaneous velocity by setting

$$q_{n+1} - q_n = \epsilon \dot{q}.$$

Note, however, that many stories are not plausible. A small ϵ does not guarantee that neighboring q values are close. Like the diffusion equation, the SE can take the particle to the Andromeda galaxy in any short time (albeit with a very small amplitude), so that the paths that interfere in giving the final answer need not be (and are not) regular and physically reasonable. Paths involving velocities exceeding c must be included. This is the type of monster multiple integrals that was introduced by N. Wiener, not in quantum mechanics but in the theory of the Brownian motion.

We can make contact with the classical Lagrangian formalism as follows. The Hamiltonian can be written as the sum $H = T + V$ of a kinetic energy $T = T(p)$, where p are the canonical momenta, plus a potential energy $V(q)$. To proceed, we write

$$\exp(-iH\epsilon) = \exp(-iT\epsilon)\exp(-iV\epsilon) + O(\epsilon^2)$$

and then neglect $O(\epsilon^2)$ since $\epsilon \to 0$ at the end. Then, introducing a set of momentum eigenstates, we write the n-th step as

$$\langle q_{n+1}|\exp(-iT\epsilon)\exp(-iV\epsilon)|q_n\rangle = \int dp_n \langle q_{n+1}|p_n\rangle\langle p_n|\exp(-iT\epsilon)\exp(-iV\epsilon)|q_n\rangle.$$

Now, using $\langle q|p\rangle = e^{ipq}$, we let the kinetic energy act on momentum eigenstates and the potential energy act on position eigenstates. One finds that

$$\langle q_{n+1}|\exp(-iT\epsilon)\exp(-iV\epsilon)|q_n\rangle = \int dp_n e^{ip_n(q_{n+1}-q_n)} e^{-i\epsilon H(p,q)}. \qquad (15.42)$$

Putting all together,

$$A(t) = \prod_n \int \frac{dq(t_n)dp(t_n)}{2\pi} \exp\left\{i\left[(p_n(q_{n+1} - q_n) - H(p_n, q_n)\epsilon\right]\right\}. \quad (15.43)$$

Note that the product produces a summation at the exponent:

$$A(t) = \left[\prod_n \left(\int \frac{dq(t_n)dp(t_n)}{2\pi}\right)\right] \exp[i \sum_j (p_j(q_{j+1} - q_j - H(p_j, q_j)\epsilon]. \quad (15.44)$$

Introducing a path integral notation for the infinite-dimensional integration, and the velocity, with $\epsilon \to d\tau$, we obtain

$$A(t) = \int Dq\, D\left(\frac{p}{2\pi}\right) \exp[i \int_t^{t'} d\tau(p\dot{q}(\tau) - H(p, q, \tau)]. \quad (15.45)$$

Under some conditions one can work out the momentum integrals.

In particular, if the Hamiltonian is $H = \frac{p^2}{2m} + V(q)$, the integrand at the exponent is of the form $\frac{p^2}{2m} + V - p\dot{q}$; then, the trick is the substitution $p \to \tilde{p} + m\dot{q}$. By performing the change of variables $\tilde{p} = p - m\dot{q}$, we are left with a simple Gaussian integral for the variable $\tilde{p}$. Actually the integrations are all of the Gaussian type $\int_{-\infty}^{\infty} dx exp(-\frac{ax^2}{2}) = \sqrt{\frac{2\pi}{a}}$, for $a > 0$.

Of course, $L(q, \dot{q}) = \frac{p^2}{2m} - V(q)$ is the Lagrangian and

$$A(t) = \int Dq \exp[i \int_t^{t'} d\tau L(q, \dot{q})]. \quad (15.46)$$

Consider two configurations $q(t_1)$ and $q(t_2)$ of a system. For each virtual path connecting them, one can calculate the action integral S. The Euler–Lagrange equations (Sect. 2) select the unique allowed classical evolution of the system. In Quantum Theory, instead, each virtual path contributes to the wave function an infinitesimal amount:

$$\Delta\Psi_{path} \propto \exp\left[\frac{2\pi i S}{h}\right]. \quad (15.47)$$

For a particle going from $\boldsymbol{x}_1(t_1)$ to $\boldsymbol{x}_2(t_2)$ in the laboratory, one must consider all trajectories, in order to get the exact result. For motions involving actions $\gg h$ the paths that are close to those that make the action stationary give a dominant contribution. In the continuum limit and in 1 dimension, the above amplitude $A(t)$ is the amplitude $P(x', t', x, t)$ that the particle is in x at time t if it is in x' at time t'. Assuming, for simplicity, a time-independent Hamiltonian, and setting $\hbar = 1$,

$$P(x', t', x, t) = \langle x'|e^{-iH(t-t')}|x\rangle. \tag{15.48}$$

The corresponding Green's function, defined as

$$G(x', t', x, t) = iP(x', t', x, t)\theta(t - t'), \tag{15.49}$$

satisfies

$$(H - i\frac{\partial}{\partial t})G = \delta(x - x')\delta(t - t'). \tag{15.50}$$

In the case of a free particle, one finds (reinserting $\hbar$)

$$P(x', t', x, t) = \left(\frac{m}{2\pi i\hbar(t - t')}\right)^{\frac{1}{2}} \exp\left(\frac{im(x - x')^2}{2\hbar(t' - t)}\right). \tag{15.51}$$

Problem 15.3 Obtain Eq. (15.51) by the path integral method.
Solution 15.3 Start directly from the above $A(t)$. All the integrals involved are Gaussian.

All this is immediately extended to 3 dimensions. Then,

$$P(\boldsymbol{x}', t', \boldsymbol{x}, t) = \left(\frac{m}{2\pi i\hbar(t - t')}\right)^{\frac{3}{2}} \exp\left(\frac{im(\boldsymbol{x} - \boldsymbol{x}')^2}{2\hbar(t' - t)}\right). \tag{15.52}$$

15.7.1 Application of the Path Integral to the Theory of the Bohm–Aharonov Effect

This formalism is well-suited to find simple approximations to problems like the diffraction by a double slit, where the dominating paths are evident (Fig. 15.1). An electron with energy E is emitted from G at time $t = 0$; the amplitude to arrive at time t_1 in slit $S1$ and hit the detector D at time T is

$$G(s1, t_1, G, 0)G(D, T, S1, t1),$$

and the total amplitude to go to point D is

$$\langle D, T|G, 0\rangle = \int_{-\infty}^{\infty} G(S1, t_1, G, 0)G(D, T, S1, t1)dt_1.$$

This integral is a convolution. Since

$$G(S1, t_1, G, 0) = \frac{i}{\hbar}\theta(t_1)(\frac{m}{2\pi i\hbar t_1})^{\frac{3}{2}} \exp(\frac{ima^2}{2\hbar t_1})$$

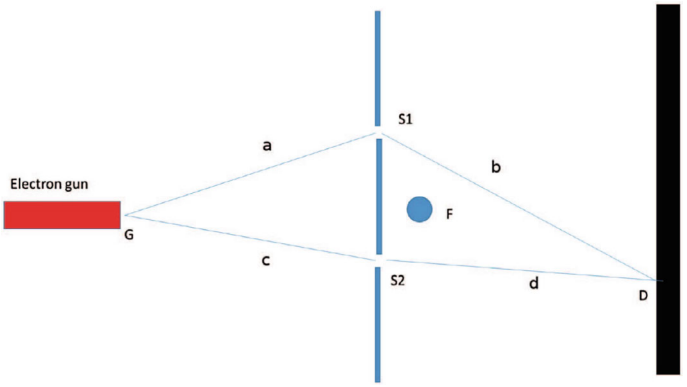


Fig. 15.1 The geometry of the double slit experiment and the Bohm–Aharonov effect

has the transform $\frac{-1}{4\pi}\frac{e^{ika}}{a}$, where $k = \sqrt{\frac{2mE}{\hbar}}$ and the transform of $G(D, T, S1, t_1)$ is $\frac{-1}{4\pi}\frac{e^{ikb}}{b}$, it is easy to transform back the product of the transforms with the result that the amplitude to go through $S1$ and arrive at time T is

$$\langle D, T|G, 0\rangle_{S1} = \frac{Z}{T^{\frac{3}{2}}}\frac{a+b}{ab}e^{\frac{im(a+b)^2}{2\hbar T}}. \tag{15.53}$$

In the same way, the amplitude to go through $S2$ is

$$\langle D, T|G, 0\rangle = \frac{Z}{T^{\frac{3}{2}}}\frac{c+d}{cd}e^{\frac{im(c+d)^2}{2\hbar T}}. \tag{15.54}$$

These amplitudes must be summed and interfere, giving us

$$\langle D, t|G, 0\rangle = |\langle D, T|G, 0\rangle_{S1} + \langle D, T|G, 0\rangle_{S2}|^2. \tag{15.55}$$

If a, b, c, d are about the same, the probability is proportional to

$$|\exp(\frac{imL_1^2}{2\hbar T}) + \exp(\frac{imL_2^2}{2\hbar T})|^2,$$

with $L_1 = a + b, L_2 = c + d$. Since $m\frac{L_1^2 - L_2^2}{T} = 2p(L_1 - L_2)$, where p is the momentum, the probability turns out to be proportional to $1 + \cos(\frac{p}{\hbar})(L_1 - L_2)$. In the presence of a magnetic flux F, the Lagrangian has an additional term $-\frac{e}{c}\boldsymbol{A}.\boldsymbol{v}$, where $\boldsymbol{v}$ is the velocity. Now,

$$\langle D, T|G, 0\rangle = \frac{Z}{T^{\frac{3}{2}}}\frac{c+d}{cd}e^{\frac{im(c+d)^2}{2\hbar T}}\exp\left(\frac{-ie}{c\hbar}\int_{C1} d\boldsymbol{x}.\boldsymbol{A}\right)\exp\left(\frac{-ie}{c\hbar}\int_{C2} d\boldsymbol{x}.\boldsymbol{A}\right) \tag{15.56}$$

where $C1$ is a path through G, S1 and D and C2 is a path through G, S2 and D. This goes like $1+\cos[\frac{m}{2\hbar T}(L_1^2-L_2^2)-\frac{e\Phi}{c\hbar}]$. Thus, the Bohm–Aharonov effect is borne out by this formalism. For the inclusion of spin in the path integral formalism, see Sect. 18.7 below.

15.8 Some Interpretations of Quantum Mechanics

In this work, I am mainly adopting the Copenhagen interpretation put forth by Bohr and Heisenberg before 1930; this is the most commonly adopted and has no shortcomings. It accepts the idea that classical and quantum worlds have different rules, and the interaction between the two worlds occurs in the process of measurement; the collapse of the wave function is governed by the instrument, which decides which aspects are measured (e.g., position or momentum of a particle, angle or angular momentum in the case of molecules and so on). Since various interpretations are adopted in the physics literature, I am presenting here the main concepts.[10] There are subtle problems, some of which could lead to observable consequences, and thus concern Science, and part of the community is not happy with the probabilistic interpretation. Albert Einstein thought that the theory is not complete and refused the use of statistical means (his motto *God does not play dice* is famous). The developments of his ideas, important results granting the coexistence with relativity and some scientific (as opposed to philosophic) aspects of the problem will be discussed below.

15.8.1 Schrödinger's Cats

Among the conceptual problems, the border between classical and quantum physics stands out, and the issue is involved. Superfluidity is a macroscopic quantum phenomenon whereby liquid He creeps up the walls of a cup. Superconductivity is a macroscopic quantum phenomenon whereby trains are levitated. White dwarfs and neutron stars are quantum objects and avoid collapse because of degeneracy and the Pauli principle (see Chap. 19). Magnetism is a purely quantum effect due to electron correlation, yet it is well known since the antiquity. So, all the macroscopic

[10] The book "Introduction to Topological Quantum Matter and Quantum Computation" by Tudor D. Stanescu presents a more detailed analysis of the main approaches to hidden variable theories.

phenomena have a deep and often quite involved quantum mechanical explanation, although a simple classical description can be enough for many purposes. Children have had magnet toys for centuries.

No classical description of microscopic phenomena is allowed, and one cannot describe atomic phenomena by treating electrons like classical marbles. An atom can be prepared in a mixed state, or superposition $\psi = c_1\psi_1 + c_2\psi_2$, where c_1, c_2 are complex numbers and ψ_1, ψ_2 orthogonal eigenfunctions of some observable $\hat{A}$ corresponding to eigenvalues a_1, a_2. A measurement of A collapses ψ, and the system must fall in one of the eigenvalues.

However, Schrödinger wished to stress that the reverse is also true, that is, it is not possible to handle macroscopic classical objects according to the microscopic rules of quantum mechanics, unless we are ready to treat the macroscopic objects in microscopic detail. The Schrödinger's cat is a well-known riddle. Suppose the cat is closed for an hour in a box with a small quantity of a radioactive element, tailored in such a way that there is a 50 percent chance that a nucleus decays in one hour. An infernal apparatus in the case of decay poisons the poor cat. Is this a way of preparing a fifty-fifty superposition of a live cat and dead cat, until the box is opened and the cat is observed? (Of course in order to assign a significance to the wave function one should prepare a large number of cats and infernal machines.)

It is obvious that the cat should be treated by classical probability. Live cat and dead cat are two classical situations, and cannot be represented as two vectors in any Hilbert space. So a question arises. Where is the border between classical and quantum? In 2012, the Nobel Prize was awarded to Serge Haroche and David J. Wineland for experimental studies in the field of Quantum Optics in which systems consisting of atoms and photons were kept 50 ms in a state of quantum superposition. In a sense, such systems are Schrödinger cats. After 50 ms, due to the interaction with the environment *decoherence* occurs, which means, a classical description prevails. Such studies pave the way to a quantum computer.

In addition, the intertwining of quantum to classical physics becomes inextricable when one considers that Quantum Mechanics treats the system according to the quantum rules but the instrument is necessarily considered classical. Philosophical or mystical interpretations have developed from this point. Some people feel that this mixing to classical physics is a difficulty. Others think that the observer has a special role in this story, yet the observer is a part of reality, which remains in the shadow. Could the cat be an observer? Famous theoreticians like Wigner proposed that the human consciousness is involved. Griffiths strongly disagrees about that. He finds that the true measurement occurs at the click of the Geiger counter and not when people look at the cat; in this sense, we can deal with the cat in classical terms.

The discussion on this topic is 90 years old and has involved many important theoreticians. Since this is my book, my personal opinion is appropriate here. This is perhaps best explained with reference to the double slit experiment (Chaps. 9 and 15). The electron wave function diffranges, and finally the particle produces a spot on a photographic plate. The diffraction is obviously a quantum effect, but the final spot is perfectly described classically. This assertion is not a way of denying the quantum nature of the chemical reactions that lead to the formation of the spot, its structure and

its persistence on the photographic plate. We describe the spot classically because the experiment shows that all those complications are quite *irrelevant when we want to describe the diffraction of the electron through the double slit*, and this agrees with the Copenhagen interpretation. In the same way, a quantum treatment of the cat is irrelevant if we just want to know if the nuclear decay occurred or not.

15.8.2 The De Broglie–Bohm Formulation of Quantum Theory

A desire for the classical trajectory and a search for a way to use it also in the quantum case has long remained among the most important theorists. A successful method is due to Louis de Broglie who originated this approach in 1927, and David Bohm, in 1952, who resumed it and developed the so-called pilot wave theory. It is an alternative interpretation of Quantum Mechanics in which a particle really obeys Classical Mechanics but is acted upon by a quantum potential. The wave function is written in the form

$$\psi = R \exp\left(\frac{i S}{\hbar}\right), \tag{15.57}$$

with $R \geq 0$ and real action S; for one particle, the action is given by the classical Hamilton–Jacobi equation with an additional non-local quantum potential

$$U = -\sum_k \frac{\hbar^2}{2m} \frac{\nabla^2 R}{R}. \tag{15.58}$$

This nonlinear framework determines particle trajectories. The theory, which has also been formulated in the many-particle case, is considered as an alternative interpretation of Quantum Theory. It gives the same physical results. It is non-local and highly nonlinear, and this is the price to pay to keep a strict analogy with Classical Mechanics. This is interesting for those who love a pilot wave description with a classical flavor, but does not lead to better understand physics; therefore I'll make it short.

15.8.3 Measurement and Wave Function Collapse

Quantum Mechanics gives no information about the way in which the interaction with the classical measurement apparatus produces the collapse, which is considered as a sort of instantaneous evolution which is outside the scope of the Scrödinger equation. For example, consider a diffraction experiment with a beam of electrons (or other charged particles or even photons) impinging on a photographic plate. A particle interacts with the plate producing a spot on the plate. The Schrödinger equation

can tell us how the spots are distributed on the plate; instead, the mechanism of the collapse of the wave function, the way the particle is absorbed and the chain of events producing the spot are not described by the wave function. The mechanism by which the apparatus produces the macroscopic signal (the spot, in this case) can be investigated separately by other experiments. The Ghirardi–Rimini–Weber approach aims to an explicit description of the collapse obtained by modifying the Hamiltonian, and/or introducing a nonlinear evolution in the course of the measurement process. The possibility that the collapse of the wave function in quantum mechanics is due to (classical) gravity has been considered, with main contributions by Diósi and Penrose. In particular, Diósi[11] proposed a noise-based dynamical reduction model, which captures the same orders of magnitude for the collapse time suggested by Penrose[12] based on heuristic arguments. In such approaches one adds a gravity-related term to the Schrödinger equation. Such proposals can be tested experimentally and the current evidence tends to be negative.[13]

Some people dislike the fact that the observer (with his consciousness) is not a part of the story. In fact it is, to the extent that the wave function depends on the experimental setup; the wave function represents what we know about the system having prepared the experiment in a particular way.

As an alternative to the collapse, Hug Everett in a thesis work in 1957 proposed a many-worlds interpretation, in which all the possible outcomes of the measurement take place in some Universe and the measurement takes us to one of these. This idea, which I find abominable, at once trivial and absurd, cannot be tested experimentally. It has influential supporters and opposers. But, while we must be ready to accept unobservable mathematical tools like vector potentials and wave functions, the description of the reality should not depend on unobservable parts of the reality itself. Science is not compatible with occultism. Thus, it appears to me that the remedy is very much worse than the disease, although it is popular among the supporters of the multiverse (an absurdity, however, which does not have really anything to do with the many-worlds interpretation, and with Science).

It is natural that the collapse of the wave function, being the transition from the quantum world to the classical one, attracts the attention of those interested in the fundamentals. It's been like this for almost a century, now. However we shall see in Chap. 27, in principle it is possible to make "weak measurements" and obtain some information without collapsing the wave function. The system can be pre-selected, weakly measured and post-selected at the end by strong measurements (by discarding all results except those corresponding to a selected final state. According to Yakir Aharonov and colleagues, this brings a new type of cat, that of Cheshire.[14] Although this effect is still questioned, the possibility of gaining information without collapsing the wave function is out of question.

[11] Diósi (1989).

[12] Penrose (1998).

[13] Vinante and Ulbricht (2021).

[14] In Lewis Carrol's book "Alice in Wonderland", Alice meets a magical cat that appears and disappears at will leaving its grin behind.

References

Diósi, L.: Phys. Rev. A **40**, 1165 (1989)
Penrose, R.: Phil. Trans. R. Soc. A **356**, 1927 (1998)
Vinante, A., Ulbricht, H.: arxiv:2109.14980pdf 30 Sep 2021
Wootters, W.K., Zurek, W.H.: A single quantum cannot be cloned. Nature **299**, 802–803 (1982)

Chapter 16
The Quantum Harmonic Oscillator

This is not just another one-dimensional example. It is a fundamental piece of the general theory.

The oscillator Hamiltonian in the coordinate representation is

$$\hat{H} = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2. \tag{16.1}$$

We know (Eq. 2.110) that by a canonical transformation it can be cast in the simpler form

$$\tilde{H} = A\omega, \tag{16.2}$$

where A is the amplitude of the oscillation. In Classical Mechanics, the amplitude of the oscillation is arbitrary, but in Quantum Mechanics, there is a natural **scale** of energies given by $E \sim \hbar\omega$. Setting

$$m\omega^2 x_0^2 \sim \hbar\omega,$$

we realize that there is also a characteristic length

$$x_0 = \sqrt{\frac{\hbar}{m\omega}}. \tag{16.3}$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_16

16.1 From the Coordinate Representation to the Second Quantization Operators

At the moment, it is not clear how to deal with the form (16.2) but from (16.1) the Schrödinger Equation (S.E.) for the stationary states $\hat{H}\psi(x) = E\psi(x)$ gives us

$$\frac{d^2}{dx^2}\psi = \frac{m^2\omega^2}{\hbar^2}x^2\psi - \frac{2mE}{\hbar^2}\psi, \tag{16.4}$$

that is, introducing the characteristic length (16.3),

$$\frac{d^2}{dx^2}\psi = \frac{x^2}{x_0^4}\psi - \frac{2m\,\mathrm{E}}{\hbar^2}\psi.$$

Introducing the dimensionless length q by $x = x_0 q$,

$$\frac{1}{x_0^2}\frac{d^2}{dq^2}\psi = \frac{q^2}{x_0^2}\psi - \frac{2mE}{\hbar^2}\psi, \tag{16.5}$$

the S.E. becomes

$$\boxed{\frac{d^2\psi}{dq^2} = (q^2 - 2\epsilon)\psi,} \tag{16.6}$$

where

$$\epsilon = \frac{E}{\hbar\omega} \tag{16.7}$$

is the dimensionless energy. Even in the Hamiltonian (16.1) one can do a similar simplification. We eliminate m in favor of x_0, by $m = \frac{\hbar}{\omega x_0^2}$, thus

$$\hat{H} = \frac{\hbar\omega}{2}\left[-\frac{d^2}{dq^2} + q^2\right], \tag{16.8}$$

which agrees with (16.6). By inspection,

$$\psi_0(q) = \frac{1}{\sqrt[4]{\pi}}e^{-\frac{q^2}{2}} \tag{16.9}$$

is a solution of the S.E. and is normalized, i.e., $\int_{-\infty}^{\infty}\psi(x)^2dx = 1$. In terms of x,

$$\psi_0(x) = Ce^{-\frac{x^2}{2x_0^2}},$$

with

$$C = \left(\frac{m\omega}{\pi\hbar}\right)^{\frac{1}{4}}$$

up to a phase factor.

This simple solution is the ground state, as we shall see. To find the general solution, we set

$$\psi(q) = h(q)e^{-\frac{q^2}{2}};$$

then, by substituting in (16.6) we find

$$h'' - 2qh' + (2\epsilon - 1)h(q) = 0. \tag{16.10}$$

By inspection, $h(q) \propto q$ is a solution with $\epsilon = \frac{3}{2}$, and one can verify easily that $h(q) \propto 4q^2 - 2$ is another solution, with $\epsilon = \frac{5}{2}$. This suggests that a systematic search for polynomial solutions is in order.

Polynomial Solutions

Setting

$$h(q) = a_0 + a_1 q + a_2 q^2 + a_3 q^3 + \cdots + a_n q^n = \sum_{j=0}^{n} a_j q^j,$$

and putting into (16.10) we get the *recurrence relation*:

$$(j+1)(j+2)a_{j+2} - 2ja_j + (2\epsilon - 1)a_j = 0, \quad j = 0, \ldots, n-2.$$

Hence, by

$$a_{j+2} = \frac{2j + 1 - 2\epsilon}{(j+1)(j+2)} a_j \quad j = 0, \ldots, n-2, \tag{16.11}$$

we can find all the even terms from a_0 and all the odd ones from a_1. To systematically find all the polynomial solutions, let us put in (16.11) the condition that $a_{N+2} = 0$; this gives us a solution of degree n.

The condition $2n + 1 - 2\epsilon_n = 0$ with (16.7) yields the eigenvalues[1]

$$\boxed{E_n = (n + \tfrac{1}{2})\hbar\omega.} \tag{16.12}$$

[1] The recurrence formula (16.11) also gives us a transcendental solution, but this is not acceptable for a wave function. For $j \to \infty$, (16.11) becomes $a_{j+2} \sim \frac{2}{j} a_j$; this is solved by $a_j \sim \frac{C}{(\frac{j}{2})!}$, with some constant C. In the even case, $h = \sum_{j=2k} \frac{C}{(\frac{j}{2})!} q^j = C \sum_k \frac{1}{k!} q^{2k} = Ce^{q^2}$; this asymptotic behavior at large q leads to $\psi \to \infty$ for $x \to \infty$, and such a solution cannot be normalized. In the odd case, $h = \sum_{j=2k-1} \frac{C}{(\frac{j}{2})!} q^j = C \sum_k \frac{1}{(k-\frac{1}{2})!} q^{2k-1}$, and since $h > C \sum_k \frac{1}{k!} q^{2k-1} \sim \frac{C}{q} e^{q^2}$, even this solution cannot be normalized.

The Gaussian solution deduced above was the ground state $N = 0$; the lowest possible energy is called zero-point energy; classically, the oscillator could be fixed at the origin, but this would violate the uncertainty principle. The characteristic energy $\hbar\omega$ is the difference between consecutive energy levels.

Hermite Polynomials

Setting $2n + 1 - 2\epsilon = 0$ in (16.10), we find the *Hermite equation*

$$H_n''(q) - 2qh_n' + 2nH_n(q) = 0; \tag{16.13}$$

the (16.11) relations become

$$a_{j+2} = \frac{2j - 2N}{(j+1)(j+2)} a_j, \quad j = 0, \ldots, N-2,$$

and the solutions are Hermite polynomials $h_N(q)$; these are orthogonal in the sense that

$$\int_{-\infty}^{\infty} H_m(x) H_n(x) e^{-x^2} dx = 0, \ m \neq n.$$

The first few are $H_0 = 1$, $H_1(q) = 2q$, $H_2(q) = 4q^2 - 2$, $H_3(q) = 8q^3 - 12q$, $H_4(q) = 16q^4 - 48q^2 + 12$.

I report without proof the following interesting formulas:

$$\frac{dH_n(x)}{dx} = 2nH_{n-1}(x) \tag{16.14}$$

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x). \tag{16.15}$$

In addition, the Rodrigues formula holds:

$$H_n(x) = (-1)^n \exp(x^2) \left(\frac{d}{dx}\right)^n \exp(-x^2). \tag{16.16}$$

This formula is useful to work out the normalization condition. First, one integrates by parts (16.1)

$$\int_{-\infty}^{\infty} e^{-x^2} H_n(x)^2 dx = \int_{-\infty}^{\infty} e^{-x^2} \frac{d^n H_n(x)}{dx^n} dx.$$

Then from (16.14) one obtains that $\frac{d^n H_n(x)}{dx^n} = 2^n n!$ So, the normalized wave function of the nth level of the harmonic oscillator is

$$\psi_n(x) = \sqrt[4]{\frac{m\omega}{\pi\hbar}} \frac{\exp[-\frac{m\omega x^2}{2\hbar}]}{\sqrt{2^n n!}} H_n\left(x\sqrt{\frac{m\omega}{\hbar}}\right). \tag{16.17}$$

16.2 Number Representation

It is important to solve again the quantum oscillator by the neat and powerful *operator method* which is basic in field theory. One would be tempted to factor the oscillator Hamiltonian (16.8) with $q = \frac{x}{x_0}$,

$$H = \frac{\hbar\omega}{2}\left[-\frac{d^2}{dq^2} + q^2\right] \stackrel{?}{=} \frac{\hbar\omega}{2}\left(\frac{d}{dq} + q\right)\left(-\frac{d}{dq} + q\right), \tag{16.18}$$

but this is wrong! The trouble is that $\frac{d}{dq}$ and q do not commute, since

$$\left[\frac{d}{dq}, q\right]_- = 1.$$

However, we can fix it. Introducing the annihilation operator

$$a = \frac{1}{\sqrt{2}}\left(q + \frac{d}{dq}\right), \tag{16.19}$$

and its Hermitian conjugate, called creation operator

$$a^\dagger = \frac{1}{\sqrt{2}}\left(q - \frac{d}{dq}\right), \tag{16.20}$$

we obtain

$$H = \hbar\omega\left(a^\dagger a + \frac{1}{2}\right) = \hbar\omega\left(\hat{n} + \frac{1}{2}\right). \tag{16.21}$$

The operator $\hat{n} = a^\dagger a$ is called number operator and its eigenvalues n label the oscillator states. The creation and annihilation operators do not commute, since

$$\boxed{[a, a^\dagger]_- = 1.} \tag{16.22}$$

Now the Schrödinger equation for the nth level, measuring energies in units $\hbar\omega$, takes the form

$$\left(\hat{n} + \frac{1}{2}\right)\psi_n \equiv \left(a^\dagger a + \frac{1}{2}\right)\psi_n = (n + \frac{1}{2})\psi_n. \tag{16.23}$$

If ψ is a solution, $a^\dagger\psi$ is also a solution, with eigenvalue $n + 1$, since the commutation rule yields $a^\dagger a a^\dagger = a^\dagger a^\dagger a + a^\dagger$, and so

$$\left(a^\dagger a + \frac{1}{2}\right) a^\dagger \psi = \left(a^\dagger [1 + a^\dagger a] + a^\dagger \frac{1}{2}\right) \psi$$

$$= a^\dagger (a^\dagger a + 1/2 + 1)\psi = a^\dagger (n+1)\psi = (n+1) a^\dagger \psi.$$

Moreover, the ground state is given by $a\psi_0 = 0$. We can verify that this is the same ground state as the one in Eq. (16.49). Indeed, in x_0 units,

$$0 = \left(x + i\frac{p}{\hbar}\right)\psi_0 = \left(x + \frac{d}{dx}\right)\psi_0. \tag{16.24}$$

Integrating, we recover $\psi_0 = e^{-\frac{x^2}{2}}$ (to be normalized).

But if ψ is any excited state, with energy eigenvalue ϵ, $a\psi$ is proportional to the solution, with energy eigenvalue $\epsilon - 1$; to see that, in

$$\left(a^\dagger a + \frac{1}{2}\right) a\psi = \left(a^\dagger a a + \frac{1}{2} a\right)\psi, \tag{16.25}$$

we put a in evidence on the left with the help of the commutation rule:

$$= \left(a a^\dagger a - a + \frac{1}{2} a\right)\psi = a\left(a^\dagger a + \frac{1}{2} - 1\right)\psi = (n-1) a\psi. \tag{16.26}$$

Evidently, $a^\dagger$ creates excitations and a destroys them; in field theories like Quantum Electrodynamics (QED) the oscillator is not observable at all. We observe, create and absorb the excitations, which are photons! Indeed, as we shall see in Chap. 19, the present oscillator commutation rules are fundamental for the description of all the Bose Fields. No excitations can be destroyed in the ground state.

In the original variables,

$$\begin{aligned} a^\dagger &= \tfrac{1}{\sqrt{2}}\left[\tfrac{x}{x_0} - \tfrac{i x_0 p}{\hbar}\right], \\ a &= \tfrac{1}{\sqrt{2}}\left[\tfrac{x}{x_0} + \tfrac{i x_0 p}{\hbar}\right]. \end{aligned} \tag{16.27}$$

The inverse transformation is

$$\begin{aligned} x &= \tfrac{x_0}{\sqrt{2}}(a + a^\dagger), \\ p &= \tfrac{-i\hbar}{x_0\sqrt{2}}(a - a^\dagger). \end{aligned} \tag{16.28}$$

One can check that $[p, x]_- = -i\hbar$.

All the eigenstates can be determined in this formalism by

$$a^\dagger \psi_n = u_n \psi_{n+1},\ a\psi_n = v_n \psi_{n-1},$$

but we must still find the constants u_n, v_n. To this end, we write the normalization condition $\langle \psi_{n+1}|\psi_{n+1}\rangle = 1$; we obtain $u_n^2 = \langle a^\dagger \psi_n | a^\dagger \psi_n\rangle = \langle \psi_n | a a^\dagger \psi_n\rangle = (1+n)$. So,

$$\psi_{n+1} = \frac{1}{\sqrt{n+1}} a^\dagger \psi_n, \tag{16.29}$$

that is,

$$|\psi_n\rangle = \frac{1}{\sqrt{n!}} (a^\dagger)^n |\psi_0\rangle. \tag{16.30}$$

Similarly, setting $a\psi_n = v_n \psi_{n-1}$, we get from $v_n^2 = \langle\, a\psi_n | a\psi_n\rangle$ that

$$a\psi_n = \sqrt{n}\psi_{n-1}. \tag{16.31}$$

Thus, $\hat{x}\psi_n = \frac{x_0}{\sqrt{2}}(\sqrt{n}\psi_{n-1} + \sqrt{n+1}\psi_{n+1})$ and the matrix of $\hat{x}$ has elements

$$x_{mn} = \frac{x_0}{\sqrt{2}}(\sqrt{n}\delta_{m,n-1} + \sqrt{n+1}\delta_{m,n+1}). \tag{16.32}$$

Similarly,

$$p_{mn} = i\hbar \frac{1}{x_0\sqrt{2}}(-\sqrt{n}\delta_{m,n-1} + \sqrt{n+1}\delta_{m,n+1}).$$

We know that

$$([\hat{p}, \hat{x}])_{mn} = (-i\hbar)_{mn} = -i\hbar\delta_{mn}.$$

We can verify that, through matrix multiplication, using the Einstein convention (sum over repeated indices),

$$([\hat{p}, \hat{x}])_{mn} \equiv (px - xp)_{mn} \equiv p_{mk}x_{kn} - x_{mk}p_{kn} = -i\hbar\delta_{mn}. \tag{16.33}$$

Similarly, one can build the matrices of $\hat{x}^2, \hat{p}^2$ and the Hamiltonian in diagonal form:

$$H_{mn} = \left(\hat{n} + \frac{1}{2}\right)\hbar\omega\delta_{mn}. \tag{16.34}$$

16.2.1 Coherent States

It is now a simple exercise to find the eigenstates of the non-Hermitian annihilation operator. Using $a|n\rangle = \sqrt{n}|n-1\rangle$, (Eq. (16.31)), one finds that

$$a|\psi_\lambda\rangle = \lambda|\psi_\lambda\rangle \tag{16.35}$$

is solved by the normalized state

$$|\psi_\lambda\rangle = \exp\left(-\frac{\lambda}{2}\right)\exp\left[\lambda a^\dagger\right]|0\rangle = \exp\left(-\frac{\lambda}{2}\right)\sum_{n=0}^{\infty}\frac{(\lambda a^\dagger)^n}{n!}|0\rangle. \tag{16.36}$$

Letting the creator operators act, we find

$$|\psi_\lambda\rangle = \exp\left(-\frac{\lambda}{2}\right)\sum_{n=0}^{\infty}\frac{\lambda^n}{\sqrt{n!}}|n\rangle. \tag{16.37}$$

Actually ψ_λ is called a *coherent state* and is very interesting for many purposes[2]. To start with, Eq. (16.35) has a useful interpretation in terms of fields. Indeed, $|\psi_\lambda\rangle$ represents a field of particles such that when a particle is removed, the field is reduced in intensity but otherwise remains the same. We shall find several important applications of this.

Moreover, suppose we replace the annihilation operator a by $s = a + \lambda$, where λ is a c-number (that is, not an operator). The consequence is that $s + s^\dagger = \sqrt{2}(q + \frac{\lambda+\lambda^*}{\sqrt{2}})$, that is, $\frac{\lambda+\lambda^*}{\sqrt{2}}$ is a shift of the origin; but since $[s, s^\dagger]_- = 1$, the new operators are canonically equivalent to the old ones. The canonically conjugated Hamiltonian is

$$\tilde{H} = \hbar\omega\left[s^\dagger s - \lambda^* s - \lambda s^\dagger + \lambda^2 + \frac{1}{2}\right] \tag{16.38}$$

and obviously has the same eigenvalues as H. However, the ground state $\tilde{\psi}$ of $\tilde{H}$ is not found by $s\tilde{\psi} = 0$ but by $(s-\lambda)\tilde{\psi} = 0$. Thus the ground state (i.e., vacuum) is a coherent state if the Hamiltonian undergoes a shift.

Problem 16.29 An oscillator is prepared in a state proportional to $|\varphi\rangle = (a^\dagger + a^{\dagger 2})|0\rangle$. Calculate $\langle x\rangle$.

Solution 16.29 The normalized state is $|\phi\rangle = \frac{1}{\sqrt{3}}(a^\dagger + a^{\dagger 2})|0\rangle$. Therefore, $\langle\phi|\hat{x}|\phi\rangle = \frac{1}{3}\frac{x_0}{\sqrt{2}}\langle 0|(a+a^2)(a+a^\dagger)(a^\dagger + (a^\dagger)^2)|0\rangle = \frac{4}{3}\frac{x_0}{\sqrt{2}}$.

The coherent state ψ_λ has further points of interest. The wave packet (16.37) may be rewritten more explicitly in terms of the harmonic oscillator wave functions as

[2] See also Chap. 25.

$$\psi_\lambda(x,t) = \exp\left(-\frac{\lambda}{2}\right)\sum_{n=0}^{\infty}\frac{\lambda^n}{\sqrt{n!}}\psi_n(x,t); \tag{16.39}$$

and the sum can be worked out thanks to the identity

$$\sum_{n=0}^{\infty} H_n(z)\frac{t^n}{n!} = \exp(z^2-(x-t)^2). \tag{16.40}$$

One obtains

$$\psi_\lambda(x,t) = \frac{1}{\sqrt{x_0\sqrt{\pi}}}\exp\left(-\frac{1}{2}\left(\frac{x}{x_0-\lambda\sqrt{2}}\right)^2\right)\exp\left(-\frac{i}{2}\varphi(t)\right), \tag{16.41}$$

where $\varphi(t)$ is a real phase. Thus,

$$|\psi(x,t)|^2 = \frac{1}{x_0\sqrt{\pi}}\exp\left[-\left(\frac{x-x_0\lambda\sqrt{2}\cos(\omega_0 t)}{x_0}\right)^2\right] \tag{16.42}$$

keeps its shape during the evolution and oscillates in a way resembling a classical oscillator. That is the reason why this is called a coherent state: it keeps oscillating in analogy with the classical oscillator. In the next section it will allow for a quantum field that is as close as possible to its classical approximation. The main source of coherent light is the Laser (see Sect. 25.4.1). Moreover, coherent states are involved in the theory of superconductivity.

16.3 Quantization of the Electromagnetic Field

There are two transverse modes of the electromagnetic field in a box (polarization along x or y, or left and right circular polarization). In the Coulomb gauge $\phi = 0, div\boldsymbol{A} = 0$, the vector potential corresponding to wave vector $\boldsymbol{k}$ and frequency ω_k is $\boldsymbol{u_{k\alpha}}(r)e^{-i\omega_k t} + c.c.$, where $u_{\boldsymbol{k\alpha}}(r)$ is normalized solution of

$$\left(\nabla^2 + \frac{\omega^2}{c^2}\right)u_{\boldsymbol{k\alpha}}(r) = 0$$

(i.e., the wave equation) and $\nabla u_{\boldsymbol{k\alpha}}(r) = 0$ (the Coulomb gauge). Consider the electric field amplitude of any monochromatic solution. It can be written as

$$E(t) = \frac{a(t)+a^*(t)}{2}, \tag{16.43}$$

where the complex amplitude $a(t)$ is proportional to $\exp(-i\omega t)$. We can treat the monochromatic solution as a harmonic oscillator by setting $q = \frac{a(0)+a^*(0)}{2}$, $p = \frac{a(0)-a^*(0)}{2i}$; so, $E(t) = q\cos(\omega t) + p\sin(\omega t)$. Here, q and p are the so-called quadrature components of the field. We have rewritten the monochromatic solution in terms of a harmonic oscillator that oscillates in amplitude space rather than in real space and has no elastic constant and no mass. Up to now the treatment is classical, but we are ready to quantize the field by imposing $[a, a^\dagger]_- = 1$. The quantum expression is of the form

$$\boldsymbol{A}_\alpha(\boldsymbol{r}, t) = \sum_k \sqrt{\frac{\hbar}{2\omega_k \epsilon_0}} [a_{k\alpha} \boldsymbol{u}_{k\alpha}(\boldsymbol{r}) e^{-i\omega_k t} + a^\dagger_{k\alpha} \boldsymbol{u}^*_{k\alpha}(\boldsymbol{r}) e^{i\omega_k t}], \tag{16.44}$$

where ϵ_0 is the electric permittivity of free space and $a^\dagger_{k\alpha}$, $a_{k\alpha}$ creation and annihilation operators of free space, which merely replace dimensionless amplitudes in what would be a classical expression; the pre-factor is chosen in such a way that the energy $E = \int [\frac{\epsilon_0}{2} \boldsymbol{E}^2 + \frac{\boldsymbol{B}^2}{2\mu_0}] d^3r$ of a quantum is $\hbar\omega$, as it should be. Note that the oscillation amplitude of the field corresponding to the x_0 of the mechanical oscillator is thereby specified. Here μ_0 is the magnetic susceptivity and $\boldsymbol{B}$ the magnetic field; the magnetic term is equal to the electric one in vacuo.

Many consequences are relevant to the appreciation of Quantum Mechanics, however. The corresponding electric field is obtained by $\mathrm{E} = -\frac{1}{c} \frac{\partial \boldsymbol{A}}{\partial t}$,

$$\boldsymbol{E}(\boldsymbol{r}, t) = \sum_{k\alpha} \sqrt{\frac{\hbar\omega_k}{2\epsilon_0}} [a_{k\alpha} \boldsymbol{u}_{k\alpha}(\boldsymbol{r}) e^{-i\omega_k t} - a^\dagger_{k\alpha} \boldsymbol{u}^*_{k\alpha}(\boldsymbol{r}) e^{i\omega_k t}]. \tag{16.45}$$

First of all, we see that the field is an operator, as every observable should be.

Moreover, it is evident that the vacuum average of each $\boldsymbol{k}$ component $\boldsymbol{E}(\boldsymbol{k}) = \sqrt{\frac{\hbar\omega_k}{2\epsilon_0}} [a_k \boldsymbol{u}_k(\boldsymbol{r}) e^{-i\omega_k t} - a^\dagger_k \boldsymbol{u}^*_k(\boldsymbol{r}) e^{i\omega_k t}]$ vanishes, but $\boldsymbol{E}(\boldsymbol{k})^2$ does not. Physically this means that if a field component is measured it is found to oscillate around zero.

The above treatment includes only the transverse modes of the electromagnetic field and ignores the rest, that is electrostatic and magnetostatic fields. Therefore it is enough for Quantum Optics, but not for Quantum Electrodynamics (QED), in which charged particles interact via the full electromagnetic interaction. In QED one adds longitudinal and time-like photons, so there are four modes. However, there are formal complications for which the reader can consult books on QED, or Advanced Quantum Mechanics by J. Sakurai.

16.3.1 Zero-Point Energy and Casimir Effect

The physical vacuum contains the zero-point oscillation and the zero-point energy of all the infinite normal modes of the electromagnetic field (and of any other physical field). The quantum vacuum is full of stuff in turmoil, and in any cube centimeter, there is infinite energy. This sounds amazing, but does not cause any difficulty; the infinite energy difference it involves is between the physical vacuum and an absolute void that does not exist in Nature. Then, one could believe that this is just formal. However, the zero-point fields do exist in the vacuum. While the average electric field is indeed zero, the average square electric field can be measured. These vacuum fields produce interesting consequences. The Lamb Shift (see next chapter) measures the effect on the atomic energy levels. Another important consequence is known as the Casimir effect. While we cannot eliminate the zero-point fields, we can change their boundary conditions, e.g., by placing two mirrors facing each other at a small distance. This produces an attractive potential that goes with the inverse fourth power of the distance between the mirrors.

Consider a metal pillbox with reflecting walls; let the basis be square with side L and the thickness be s. The modes with $k = (\frac{\pi a}{s}, \frac{\pi b}{s}, \frac{\pi c}{s})$ with integer a, b and c contribute $h\nu_k = \hbar ck$ to the energy of the vacuum (indeed, they contribute twice because of the two polarizations). The energy of the vacuum fields in the pillbox, $U(s) = \sum_{modes} \hbar ck$ diverges, but we are interested in a simple estimate of the effect of the presence of the walls. Any free electron metal is characterized by a number density $n(\mathbf{r})$ of conduction electrons capable of screening external fields, and by a plasma frequency

$$\omega_p = \sqrt{\frac{4\pi ne^2}{m}},$$

where n is the number density of conduction electrons and m is their effective mass. As an order of magnitude, we can safely assume $\hbar\omega_p \sim 10$ eV. The electrons are only able to screen fields below ω_p and do not discriminate between zero point and applied fields.

A model calculation can hardly be realistic, but it can show the finiteness of the effect and its s dependence. Using a cutoff length α to avoid the divergence,

$$U(s) = \pi\hbar c \sum_{a,b,c}^{\infty} \lambda(a,b,c) \exp(-\alpha\lambda(a,b,c)), \tag{16.46}$$

where $\lambda(a,b,c) = \sqrt{\frac{a^2}{s^2} + \frac{b^2}{L^2} + \frac{c^2}{L^2}}$ is an inverse length. One wishes eventually to remove the cutoff with $\alpha \to 0$. It is evident that the result must be proportional to the area L^2 and for dimensional reasons. The summation gives us

$$U(s) = \frac{\hbar c\pi^2 L^2}{2} \left(\frac{d}{d\alpha}\right)^2 \left(\frac{1}{\alpha}\frac{1}{e^{\frac{\alpha}{s}}} - 1\right). \tag{16.47}$$

Now, if a cavity with a width W is divided into two halves by a mirror, the vacuum energy is $2U(\frac{W}{2})$.Instead, if the mirror is at a distance s from one wall and $W-s$ from the other, the vacuum energy is $U(s)+U(W-s)$. All these energies diverge, but the energy cost of shifting the mirror from a distance s to the middle of the cavity is finite. It is the Casimir energy

$$C(s) = U(s) + U(W-s) - 2U(\frac{W}{2}) \sim \frac{\pi^2 \hbar c}{720 s^3}. \tag{16.48}$$

More precisely, the attractive force per unit area ($1\,\mathrm{cm}^2$) between perfect conductors at distance s is calculated to be $\frac{0.013}{s^4}$ dyne, and the effect has been confirmed experimentally in 1997 by S. Lamoreaux.

The numerical values in the above derivation cannot be taken too seriously. By varying the nature of the surfaces and angles, one can control the Casimir force, which, in some cases, can also be repulsive, and perhaps one day, this will be used to operate nano-engines. The vacuum energy due to all possible fields is also suspected to be at the origin of the Λ term which produces the dark energy (Sect. 8.12), but no quantitative theory is available to get from the Planck length $\lambda_P = \sqrt{\frac{\hbar G}{c^3}} \sim 1.6 \quad 10^{-35}\,\mathrm{m}$ and the Planck mass $M_P \sim 2.2 \quad 10^{-8}\,\mathrm{kg}$ to the estimated energy density of dark energy $10^{-29}\,\frac{g}{\mathrm{cm}^3}$.

Other effects of the vacuum fields will be discussed at the end of the next chapter.

16.3.2 Parametric Down Conversion and Squeezed Light

In the above discussion of the harmonic oscillator, I have emphasized that in the quantum case, there is a fundamental length x_0, which characterizes the eigenstate wave functions in both real and momentum space. However, it is possible to manipulate the oscillator by putting it in a nonlinear medium in such a way that the wave functions are scaled. For instance, the wave function of the squeezed vacuum with squeezing parameter R is obtained from the oscillator ground state (16.9) by scaling

$$\psi_R(q) = \frac{\sqrt{R}}{\sqrt[4]{\pi}} e^{-\frac{(Rq)^2}{2}}. \tag{16.49}$$

It is, of course, a normalized solution, i.e., $\int_{-\infty}^{\infty} \psi_R(x)^2 dx = 1$. The scaling modifies the variances, with $\Delta q^2 = \frac{1}{2R^2}$, $\Delta p^2 = \frac{R^2}{2}$, in such a way that the uncertainty in one of the quadratures (x or p) is decreased but the uncertainty in the other quadrature is increased so that the Heisenberg uncertainty principle is preserved.

It is possible to squeeze the position of an oscillator by producing the state $|\psi\rangle = |0\rangle - \frac{s}{\sqrt{2}}|2\rangle$. In this state, $\langle\psi|x^2|\psi\rangle = \frac{1}{2} - s$, so the spread in x is reduced.

With two modes a and b the vacuum

$$\Psi_{00} = \frac{1}{\sqrt{\pi}} \exp\left(-\frac{q_a^2 + q_b^2}{2}\right) = \frac{1}{\sqrt{\pi}} \exp\left(-\frac{(q_a + q_b)^2}{4}\right) \exp\left(-\frac{(q_a - q_b)^2}{4}\right)$$

can be squeezed to become

$$\Psi_{00sq} = \frac{1}{\sqrt{\pi}} \exp\left(-\frac{(q_a + q_b)^2}{4R^2}\right) \exp\left(-\frac{R^2(q_a - q_b)^2}{4}\right),$$

where R is called the squeezing parameter. One way to do that is by using Parametric Down Conversion. This is a nonlinear optical phenomenon occurring in crystals like beta Barium Borate under intense laser light. A fraction $f \ll 1$ of the photons is split into pairs of photons having half frequency, and thus half of the energy and momentum each. The two photons may have the same polarization (Type I correlation) or perpendicular polarizations (Type II). Let us consider the latter case. The half frequency mode oscillators are no longer in the vacuum state but in a state

$$|S\rangle = |00\rangle + f|11\rangle.$$

Let a and b denote the two polarizations and q_a and q_b denote the dimensionless coordinates of the oscillators $q_a = \frac{a+a^\dagger}{\sqrt{2}}$, $q_b = \frac{b+b^\dagger}{\sqrt{2}}$. The variance of $q_a - q_b$ is given by $\langle S|(q_a - q_b)^2|S\rangle$. Expanding, one finds that the only terms that contribute are those in $aa^\dagger + a^\dagger a + bb^\dagger + b^\dagger b - 2ab - 2a^\dagger b^\dagger$ and the variance turns out to be $1 - 2f + O(f^2)$.

Similarly, the variance of $p_a - p_b$ is given by $\langle S|(p_a - p_b)^2|S\rangle = 1 + 2f + O(f^2)$, as one can verify by setting $p_a = \frac{-i(a-a^\dagger)}{\sqrt{2}}$, $p_b = \frac{-i(b-b^\dagger)}{\sqrt{2}}$ and working out the calculation in the same way. This calculation implies an increased correlation between the coordinates and a decreased correlation between the momenta. This implies that the wave function of the two mode vacuum state has become a *squeezed vacuum*.

Squeezed coherent states have also been obtained, e.g., by nonlinear optics techniques. Squeezed states have important applications in precision measurements (clocks, interferometers) and in gravitational wave detectors.

Chapter 17
Stationary States of One Particle in 3 Dimensions

The partial differential equations are quite a bit harder to solve than the ordinary ones, unless the symmetry allows us to separate the variables. Fortunately, among the most interesting stationary problems, there are some that can be solved analytically.

17.1 Separation of Variables in Cartesian Coordinates

The three-dimensional plane wave is the product of one-dimensional plane waves, and the kinetic energy is the sum of the contributions of motions along x, y, z. More generally, the problem is separable into Cartesian coordinates if the potential energy is of the form

$$V(x, y, z) = U_x(x) + U_y(y) + U_z(z) \tag{17.1}$$

where $U_x(x)$, $U_y(y)$ and $U_z(z)$ are arbitrary functions. In this case, the motions along the three directions are independent. In the Schrödinger stationary state equation

$$\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right]\Psi(x, y, z) + V(x, y, z)\Psi(x, y, z) = E\Psi(x, y, z),$$

one puts the factored solution

$$\Psi(x, y, z) = X(x)Y(y)Z(z); \tag{17.2}$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_17

dividing by $\Psi(x, y, z)$,

$$\frac{1}{X(x)}\frac{\partial^2 X}{\partial x^2} + U_x(x) + \frac{1}{Y(y)}\frac{\partial^2 Y}{\partial y^2} + U_z(z) + \frac{1}{Z(z)}\frac{\partial^2 Z}{\partial z^2} + U_z(z) = E.$$

Therefore,

$$\begin{cases} \dfrac{1}{X(x)}\dfrac{\partial^2 X}{\partial x^2} + U_x(x) = \epsilon_x \Rightarrow \dfrac{\partial^2 X}{\partial x^2} + U_x(x)X(x) = \epsilon_x X(x), \\ \dfrac{1}{Y(y)}\dfrac{\partial^2 Y}{\partial y^2} + U_y(y) = \epsilon_y \Rightarrow \dfrac{\partial^2 Y}{\partial y^2} + U_y(y)Y(y) = \epsilon_y Y(y), \\ \dfrac{1}{Z(z)}\dfrac{\partial^2 Z}{\partial z^2} + U_z(z) = \epsilon_z \Rightarrow \dfrac{\partial^2 Z}{\partial z^2} + U_z(z)Z(z) = \epsilon_z Z(z), \\ \epsilon_x + \epsilon_y + \epsilon_z = E, \end{cases}$$

and the problem is broken. In this way, one gets special solutions, but since they form a complete set (Second Postulate) the most general solution can be expanded as a convergent series of those special solutions.

Box

Consider a particle confined in a potential

$$V(x, y, z) = \begin{cases} 0, -\frac{L_x}{2} < x < \frac{L_x}{2}, -\frac{L_y}{2} < y < \frac{L_y}{2}, -\frac{L_z}{2} < z < \frac{L_z}{2}, \\ \infty \quad \text{otherwise.} \end{cases}$$

V is of the form (17.1). Separating the variables, one finds the eigenfunctions

$$\psi_{n_x n_y n_z}(x, y, z) = u_{n_x}(x)u_{n_y}(y)u_{n_z}(z)$$

and the energy eigenvalues

$$E_{n_x n_y n_z}(x, y, z) = \epsilon_{n_x} + \epsilon_{n_y} + \epsilon_{n_z},$$

with

$$\begin{aligned} u_{n_x} &= \sqrt{\tfrac{2}{L_x}} \sin\left[\tfrac{\pi n_x}{L_x}\left(x + \tfrac{L_x}{2}\right)\right], \\ \epsilon_{n_x} &= \hbar^2 \tfrac{\pi^2 n_x^2}{2mL_x^2}, \end{aligned}$$

etc. The motions along x, y and z are independent; while classically one multiplies the probabilities, here the amplitudes are multiplied. It is *statistical independence*.

In the cubic case $L_x = L_y = L_z = L$, many levels are *degenerate*, that is, several choices of the quantum numbers give the same energy. For instance, $E_{511} = E_{151} = E_{115} = E_{333} = 27\hbar^2 \frac{\pi^2}{2mL^2}$. This is a very important and general remark. More sym-

metry generally implies more degeneracy, when there are several noncommuting symmetry operations that commute with the Hamiltonian. This idea is fully developed by Group Theory.

Oscillator in 3 dimensions

Similarly, one solves the problem

$$H = \frac{p^2}{2m} + \frac{1}{2}m\left[\omega_x^2 x^2 + \omega_y^2 y^2 + \omega_z^2 z^2\right],$$

still with $E_{n_x n_y n_z}(x, y, z) = \epsilon_{n_x} + \epsilon_{n_y} + \epsilon_{n_z}$ and high degeneracy in the isotropic case $\omega_x = \omega_y = \omega_z = \omega$.

17.2 Landau Levels

The Hamiltonian of a charged particle in a magnetic field $\overrightarrow{H}$ is of special interest for several applications. Let us take the z axis parallel to $\overrightarrow{H}$. We know from Classical Mechanics that the Hamiltonian $\hat{H}$ depends on the vector potential $\overrightarrow{A}$:

$$\hat{H} = \frac{(\overrightarrow{p} - \frac{e}{c}\overrightarrow{A})^2}{2m}. \tag{17.3}$$

We use the first Landau[1] gauge

$$\overrightarrow{A} = (-y, 0, 0)B. \tag{17.4}$$

This does not preserve the symmetry of the problem. There is also a second Landau gauge

$$\overrightarrow{A} = (0, x, 0)B, \tag{17.5}$$

and a third Landau gauge

$$\overrightarrow{A} = (-y, x, 0)\frac{B}{2}. \tag{17.6}$$

This treats x and y on equal footing. Using the first Landau gauge,

$$\hat{H} = \frac{(p_x + \frac{eBy}{c})^2 + p_y^2 + p_z^2}{2m}. \tag{17.7}$$

[1] Lev Davidovic Landau (Baku 1908–Moscow 1968) was the most important Soviet physicist. He won the Nobel prize in 1962 for his works on superfluid He.

Here, p_x and p_z are conserved, and we can take the wave function in the form

$$\psi(x, y, z) = e^{i[p_x x + p_z z]}\chi(y); \tag{17.8}$$

along the x and z axes the particle is free, and one can replace operators with eigenvalues, $p_x \to \hbar k_x$ and $p_z \to \hbar k_z$. Thus, inserting the cyclotron frequency $\frac{eB}{m} = \omega_C$

$$\hat{H} = \frac{\hat{p_y}^2 + (\hbar k_x - eBy)^2 + (\hbar k_z)^2}{2m}. \tag{17.9}$$

Since this simply involves a shift in the origin of y, the motion along y is quantized[2] and harmonic with frequency ω_C. The energy eigenvalues are

$$E_n(k_z) = (n + \frac{1}{2})\hbar\omega_C + \frac{(\hbar k_z)^2}{2m.} \tag{17.10}$$

The quantized levels (at fixed k_z) are called *Landau levels*.

17.3 Separation of Variables in Spherical Coordinates

When the potential energy depends only on the distance from a point, which is conveniently taken as the origin (central problems) the S.E. is separable in spherical coordinates; the transformations from Cartesian to spherical coordinates and back are

$$\begin{cases} x = rsin\theta\cos\phi, \\ y = rsin\theta\sin\phi, \\ z = r\cos\theta, \end{cases} \quad \begin{cases} r = \sqrt{x^2 + y^2 + z^2}, \\ \theta = \arccos(\frac{z}{\sqrt{x^2+y^2+z^2}}), \\ \phi = \arctan(\frac{y}{x}). \end{cases} \tag{17.11}$$

By using the chain rule and elementary, lengthy algebra (which is best done by computer), one finds the momentum components

$$\begin{cases} ip_x = \sin(\theta)\cos(\phi)\dfrac{\partial}{\partial r} + \dfrac{\cos(\phi)\cos(\theta)}{r}\dfrac{\partial}{\partial\theta} - \dfrac{\sin(\phi)}{r\sin(\theta)}\dfrac{\partial}{\partial\phi} \\ ip_y = \sin(\theta)\sin(\phi)\dfrac{\partial}{\partial r} + \dfrac{\sin(\phi)\cos(\theta)}{r}\dfrac{\partial}{\partial\theta} + \dfrac{\cos(\phi)}{r\sin(\theta)}\dfrac{\partial}{\partial\phi} \\ ip_z = \cos(\theta)\dfrac{\partial}{\partial r} - \dfrac{\sin(\theta)}{r}\dfrac{\partial}{\partial\theta}. \end{cases} \tag{17.12}$$

[2] The special role of the y direction is due to the gauge; by a gauge transformation we can rotate the pair x, y around the field direction as we like. The wave function is largely arbitrary and the physical quantities do not depend on the gauge.

We know already that $L_z = xp_y - yp_x = -i\frac{\partial}{\partial\phi}$. Moreover, we find

$$L_\pm \equiv L_x \pm iL_y = e^{\pm i\phi}\left[\frac{\partial}{\partial\theta} \pm i\cot(\theta)\frac{\partial}{\partial\phi}\right]$$

and

$$-L^2 = \frac{1}{\sin(\theta)}\frac{\partial}{\partial\theta}\sin(\theta)\frac{\partial}{\partial\theta} + \frac{1}{\sin(\theta)^2}\frac{\partial^2}{\partial\phi^2}.$$

Finally, the Laplacian in spherical coordinates is

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2(\theta)}\frac{\partial^2}{\partial\phi^2}. \quad (17.13)$$

If $V = V(r)$, the stationary state equation

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(r)\right]\Psi(r,\theta,\phi) = E\Psi(r,\theta,\phi)$$

splits in a radial equation and an angular one by putting

$$\Psi(r,\theta,\phi) = R(r)Y_{lm}(\theta,\phi).$$

Let $|\theta,\phi\rangle$ denote the amplitude that the system has a sharp orientation in space in the direction with well-defined angles θ,ϕ. This may be the angular factor of a wave function of a particle which separates in spherical coordinates. The *Spherical Harmonics*

$$Y_{lm}(\theta,\phi) = \langle\theta\phi|l,m\rangle$$

satisfy the angular equation

$$\frac{1}{\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial Y}{\partial\theta}\right) + \frac{1}{\sin^2(\theta)}\frac{\partial^2 Y}{\partial\phi^2} = -\lambda Y, \quad (17.14)$$

while the *radial function* is obtained by solving

$$\frac{d}{dR}\left(r^2\frac{dR}{dr}\right) - \frac{2mr^2}{\hbar^2}\left[V(r) - E\right]R = \lambda R. \quad (17.15)$$

17.3.1 Spherical Harmonics

The angular Eq. (17.14) separates if we put

$$Y_{lm}(\theta, \phi) = \Theta_{lm}(\theta)\Phi_m(\phi), \tag{17.16}$$

where

$$\frac{d^2\Phi_m}{d\phi^2} = -m^2\Phi_m \quad \Rightarrow \Phi_m = e^{im\phi}, \text{ integer m.} \tag{17.17}$$

We have already encountered the eigenfunctions of L_z. The Θ equation, known as the Legendre equation, reads as

$$\sin(\theta)\frac{d}{d\theta}\left(\sin(\theta)\frac{d\Theta_{lm}}{d\theta}\right) + \{\lambda \sin^2(\theta) - m^2\}\Theta_{lm} = 0, \tag{17.18}$$

where we know from Sect. 13.1 that $\lambda = l(l+1)$. For $l = m = 0$, this is simply $\frac{\partial}{\partial\theta}(\sin(\theta)\frac{\partial}{\partial\theta})\Theta(\theta) = 0$ and is solved by $\Theta = 1$. For $l = 1$, $\lambda = 2$, one readily finds that for $m = 0$, $\Theta = \cos(\theta)$, and for $m = \pm 1$, $\Theta = \sin(\theta)$. In general, it is elementary to find that the solution is a polynomial in $\cos(\theta)$ for $m = 0$, otherwise it is a polynomial in $\cos(\theta)$ and $\sin/\theta)$. Indeed, the general solution is $\Theta(\theta) = P_l^m(\cos(\theta))$, where

$$P_l^m(x) = (-1)^m(1-x^2)^{\frac{m}{2}}\frac{d^m}{dx^m}P_l(x) \tag{17.19}$$

is an *associated Legendre polynomial*. The Legendre polynomials $P_l(x)$ satisfy

$$\frac{d}{dx}[(1-x^2)\frac{d}{dx}P_m(x)] + m(m+1)P_m(x) = 0. \tag{17.20}$$

In particular, $P_0(x) = 1$, $P_1(x) = x$, $P_2(x) = \frac{3x^2-1}{2}$, $P_3(x) = \frac{5x^3-3x}{2}$. They can be computed from the Rodriguez formula

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n.$$

Then, $Y_{lm}(\theta, \phi) = \langle\theta, \phi|l, m\rangle$ is the amplitude in the direction defined by θ and ϕ of an eigenfunction of L^2 with eigenvalue $\hbar^2 l(l+1)$ and L_z with eigenvalue $m\hbar$.

The spherical harmonics are tabulated in terms of associated Legendre functions, which are also easily available.

l	m	Y_{lm}	$Y_{lm}r^l$
0	0	$\frac{1}{\sqrt{4\pi}}$	$\frac{1}{\sqrt{4\pi}}$
1	0	$\sqrt{\frac{3}{4\pi}}\cos\theta$	$\sqrt{\frac{3}{4\pi}}\cos\theta z$
1	±1	$\pm\sqrt{\frac{3}{8\pi}}\sin\theta e^{\pm i\phi}$	$\pm\sqrt{\frac{3}{8\pi}}(x\pm iy)$
2	0	$\sqrt{\frac{5}{16\pi}}(3\cos^2\theta-1)$	$\sqrt{\frac{5}{16\pi}}(3z^2-r^2)$
2	±1	$\pm\sqrt{\frac{15}{8\pi}}\cos\theta\sin\theta e^{\pm i\phi}$	$\pm\sqrt{\frac{15}{8\pi}}z(x\pm iy)$
2	±2	$\sqrt{\frac{15}{32\pi}}\sin^2\theta e^{\pm 2i\phi}$	$\sqrt{\frac{15}{32\pi}}(x\pm iy)^2$

The closure relation is[3]

$$\sum_{k=0}^{\infty}\sum_{m=-k}^{k} Y^*_{km}(\theta_1,\phi_1)Y_{km}(\theta_2,\phi_2) = \delta(\Omega_1-\Omega_2) = \frac{\delta(\theta_1-\theta_2)\delta(\phi_1-\phi_2)}{|\sin(\theta_1)|}.$$

This means that any good $f(\theta,\phi)$ can be developed:

$$f(\theta,\phi) = \sum_{k=0}^{\infty}\sum_{m=-k}^{k}\int Y^*_{km}(\theta_1,\phi_1)f(\theta_1,\phi_1)d\Omega_1 Y_{km}(\theta,\phi). \tag{17.21}$$

Parity is the operation $P:(x,y,z)\to(-x,-y,-z)$. In order to change the sign of $z=r\cos(\theta)$, one can do $\theta\to\pi-\theta$, which implies $\sin(\theta)\to\sin(\theta)$; to change the sign of x and y, one can do $\phi\to\phi+\pi$. So,

$$P:\phi\to\phi+\pi,\quad \theta\to\pi-\theta.$$

Under the action of P, $Y_{ll}=e^{li\phi}(\sin^l(\theta))$ takes a factor $e^{il\pi}=(-1)^l$. This result cannot depend on m, since $L_\pm$, like the components of the pseudo-vector angular momentum, are even under parity.

17.4 Central Field

In a central potential $V(r)$, we use spherical coordinates and seek special solutions of the form

$$\psi(r,\theta,\phi) = R(r)Y_{lm}(\theta,\phi);$$

[3] Recall that

$$\delta(g(x)) = \frac{\sum_\alpha \delta(x-x_\alpha)}{\left|\frac{dg}{dx}\right|}.$$

then,

$$\begin{aligned} -\nabla^2\psi &= \left[\tfrac{-1}{r^2}\tfrac{\partial}{\partial r}\left(r^2\tfrac{\partial}{\partial r}\right) + \tfrac{L^2}{\hbar^2 r^2}\right]\psi \\ &= \left[\tfrac{-1}{r^2}\tfrac{\partial}{\partial r}\left(r^2\tfrac{\partial}{\partial r}\right) + \tfrac{l(l+1)}{r^2}\right]\psi. \end{aligned}$$

Removing the spherical harmonic, we are left with the *radial equation*

$$\frac{-1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right) + \frac{l(l+1)}{r^2}R + \frac{2mV(r)}{\hbar^2}R = k^2R(r), \tag{17.22}$$

where

$$k^2 = \frac{2mE}{\hbar^2}.$$

We gain some simplification by setting

$$R = \frac{u}{r};$$

then,

$$\frac{\hbar^2}{2m}\left[\frac{-1}{r}\frac{d^2u}{dr^2} + \frac{l(l+1)}{r^2}\frac{u}{r}\right] + V(r)\frac{u}{r} = E\frac{u}{r},$$

that is,

$$-\frac{\hbar^2}{2m}\frac{d^2u}{dr^2} + \underbrace{\frac{\hbar^2}{2m}\frac{l(l+1)}{r^2}u(r) + V(r)u(r)} = Eu(r).$$

This is a one-dimensional problem with a modified potential. The term $l(l+1)$ is a centrifugal potential and forces the wave function to vanish at the center unless $l = 0$. It is convenient to write

$$-\frac{d^2u}{dr^2} + \frac{l(l+1)}{r^2}u(r) + \frac{2m}{\hbar^2}V(r)u(r) = k^2u(r).$$

Free Particle Again!

For $V = 0$ we get

$$\frac{d^2u}{dr^2} = \frac{l(l+1)}{r^2}u(r) - k^2u(r).$$

Equation (17.22) with $\rho = kr$ is the *spherical Bessel equation*

$$2\rho R' + \rho^2 R'' + (\rho^2 - l(l+1))R = 0,$$

and the solutions are

$$j_l(\rho) = \rho^l \left(-\frac{1}{\rho}\frac{d}{d\rho}\right)^l \frac{\sin(\rho)}{\rho}.$$

The free particle eigenstates of H with diagonal L^2 and L_z are quite different from the De Broglie eigenstates of H and momentum. For $l = 0$,

$$u_{l=0}(r) = \sin(kr) \Longrightarrow R_{l=0}(r) = \frac{\sin(kr)}{r}.$$

The alternative solution $u_{l=0}(r) = \cos(kr)$ must be discarded, because $R_{l=0}(r)$ would blow up at the origin. We can expand the plane wave in spherical solutions.

This is simplest for e^{ikz}, since we need only harmonics with $m = 0$, $Y_{l0}(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi}} P_l(\cos(\theta))$, with P_l Legendre polynomials. One can show that

$$e^{ikz} = \sum_{l=0}^{\infty} (-i)^l (2l+1) P_l(\cos(\theta)) (\frac{r}{k})^l (\frac{1}{r}\frac{d}{dr})^l \frac{\sin(kr)}{r}.$$

Spherical Potential Well

By imposing boundary conditions for $r = R_0$, one finds the energy eigenvalues for a particle in a spherical potential well

$$V(r) = \begin{cases} V = 0 & r < a, \\ V = \infty & r > a. \end{cases}$$

In the case $l = 0$, one finds that

$$E_{n0} = \frac{\hbar^2 \pi^2}{2ma^2} n^2, \ n = 1, 2, 3, \ldots$$

For general l the eigenvalues E_{nl} are

$$E_{nl} = \frac{\hbar^2 x_{nl}^2}{2ma^2},$$

where x_{nl} is the n-th root of $j_l(x) = 0$ and j_l is the spherical Bessel function.

17.5 The Hydrogenoid Atom

We now solve the Schrödinger equation for the bound states of the H atom and one-electron atoms with the nuclear charge Ze. It is fair to mention, however, that these results were first obtained by Pauli[4] in a masterpiece of a paper before the formulation of the Schrödinger equation, by using the conservation of the quantum Runge–Lenz vector

$$\boldsymbol{R} = \frac{\boldsymbol{p} \wedge \boldsymbol{L} - \boldsymbol{L} \wedge \boldsymbol{p}}{2m} - k\frac{\boldsymbol{r}}{r}, \tag{17.23}$$

where k is the coefficient of $\frac{1}{r}$ in the Hamiltonian. It is modified from the classical expression (Chap. 2) in order to preserve the property that it commutes with the Hamiltonian.

For $V(r) = -\frac{Ze^2}{r}$, the radial equation reads as

$$-\frac{d^2u}{dr^2} + \frac{l(l+1)}{r^2}u - \frac{2m}{\hbar^2}\frac{Ze^2}{r}u = \frac{2mE}{\hbar^2}u.$$

The classical Kepler problem has no length scale, and the orbit can have any size. In Quantum Mechanics, it is the existence of $\hbar$ that determines the atomic radius. Since $\frac{d^2u}{dr^2}$ is an inverse square length, the characteristic size is

$$a_0 = \frac{\hbar^2}{me^2Z} = \frac{a_B}{Z},$$

where a_B is the so-called *Bohr radius*; we must solve

$$-\frac{d^2u}{dr^2} + \frac{l(l+1)}{r^2}u - \frac{2}{a_0 r}u = \frac{2mE}{\hbar^2}u.$$

It turns out that $a_B = 0.529$ Angström, where 1 Angström $=10^{-8}$ cm. It is convenient to put the equation in a dimensionless form by setting

$$\rho = \frac{r}{a_0}, \quad \epsilon = \frac{2ma_0^2E}{\hbar^2} < 0 \tag{17.24}$$

and multiplying by a_0^2. The dimensionless form is

$$u'' + \epsilon u(\rho) = [-\frac{2}{\rho} + \frac{l(l+1)}{\rho^2}]u(\rho). \tag{17.25}$$

[4] Pauli (1926).

Special Case

The case[5] $l = 0$ is simplest and for $\rho \to \infty$, by approximating by $u'' + \epsilon u(\rho) \sim 0$, one finds the trend $u \to e^{-\rho\sqrt{-\epsilon}}$: this prompts the trial solution

$$u = \rho e^{-\lambda\rho}, \ \lambda = \sqrt{-\epsilon}.$$

It turns out that

$$u' = (1 - \lambda\rho)e^{-\lambda\rho},$$
$$u'' = (-2\lambda + \lambda^2\rho)e^{-\lambda\rho} = (\tfrac{-2\lambda}{\rho} + \lambda^2)\rho e^{-\lambda\rho}.$$

So, $\epsilon = -1$. These results concern the ground state, as we shall see, and lead to

$$R(r) \propto e^{-\frac{r}{a_0}} = e^{-\frac{Zr}{a_B}}, \quad E = -\frac{\hbar^2}{2\,m a_0^2} = -\frac{1}{2}\frac{m e^4 Z^2}{\hbar^2}.$$

Note the strong Z-dependence. Moreover, $E = -\frac{Ze^2}{2a_0}$.

Normalization

Since $\frac{2}{\sqrt{a_0^3}} e^{-\frac{r}{a_0}}$ is normalized with

$$\int_0^\infty dr r^2 |R_{1,0}|^2 = 1$$

and $Y_{00} = \frac{1}{\sqrt{4\pi}}$,

$$\psi = \frac{1}{\sqrt{\pi a_0^3}} e^{-\frac{r}{a_0}}.$$

Excited Bound States

From (17.25), we see that even if $l > 0$, the asymptotic trend for $\rho \to \infty$ remains $u \to e^{-\rho\sqrt{-\epsilon}}$; the short distance behavior changes at short distances from the center because of the centrifugal barrier, and becomes

$$u'' \sim \frac{l(l+1)}{\rho^2} u(\rho), \quad \rho \to 0.$$

Therefore, $u \to \rho^{l+1}$. Then, we set
$\lambda = \sqrt{-\epsilon}$, with

$$\begin{aligned} u &= e^{-\lambda\rho}\rho^{l+1}\left[c_0 + c_1\rho + \cdots + c_{n_r}\rho^{n_r}\right] \\ &= e^{-\lambda\rho}\left[c_0\rho^{l+1} + c_1\rho^{l+2} + \cdots + c_{n_r}\rho^{l+1+n_r}\right], \end{aligned}$$

[5] The states with $l = 0, 1, 2, 3, \ldots$, are called s, p, d, f, g, h, i, etc. and so on, in alphabetic order.

that is,

$$u(\rho) = e^{-\lambda\rho} \sum_{\nu=0}^{n_r} c_\nu \rho^{\nu+l+1}. \tag{17.26}$$

Therefore, the solution is of the form

$$u = e^{-\lambda\rho} \rho^{l+1} f^{(n_r,l)}(\rho), \tag{17.27}$$

where n_r is the *radial quantum number*. We shall find that $n_r \to \infty$ cannot be accepted, and $f^{(n_r,l)}(\rho)$ is a polynomial.

Then, differentiating (17.26),

$$u' = e^{-\lambda\rho} \sum_{\nu=0}^{n_r} c_\nu [-\lambda\rho^{\nu+l+1} + (\nu + l + 1)\rho^{\nu+l}].$$

Differentiating again,

$$\begin{aligned} u'' = e^{-\lambda\rho} \sum_{\nu=0}^{n_r} c_\nu \{ & -\lambda[-\lambda\rho^{\nu+l+1} + (\nu + l + 1)\rho^{\nu+l}] \\ & -\lambda(\nu + l + 1)\rho^{\nu+l} + (\nu + l + 1)(\nu + l)\rho^{\nu+l-1}\}. \end{aligned}$$

Let us simplify

$$u'' = e^{-\lambda\rho} \sum_{\nu=0}^{n_r} c_\nu \{\lambda^2\rho^{\nu+l+1} - 2\lambda(\nu + l + 1)\rho^{\nu+l} + (\nu + l)(\nu + l + 1)\rho^{\nu+l-1}\}.$$

The first term is $\lambda^2 \sum c_\nu \rho^{\nu+l+1} e^{-\lambda\rho} = -\epsilon u$. Therefore,

$$u'' + \epsilon u = e^{-\lambda\rho} \left\{ -2\lambda \sum_{\nu=0}^{n_r} c_\nu (\nu + l + 1)\rho^{\nu+l} + \sum_{\nu=0}^{n_r} c_\nu (\nu + l)(\nu + l + 1)\rho^{\nu+l-1} \right\};$$

we shift the second sum by renaming $\nu \to \nu + 1 : \sum_{\nu=a}^{b} f(\nu) = \sum_{\nu+1=a}^{b} f(\nu + 1) = \sum_{\nu=a-1}^{b-1} f(\nu + 1)$. To avoid complications with the first and last terms, we put $c_\nu = 0$ for $\nu < 0$ and for $\nu > n_r$. So,

$$\begin{aligned} u'' + \epsilon u = e^{-\lambda\rho} \Bigg\{ & -2\lambda \sum_{\nu=0}^{n_r} c_\nu (\nu + l + 1)\rho^{\nu+l} \\ & + \sum_{\nu=-1}^{n_r-1} c_{\nu+1} (\nu + l + 1)(\nu + l + 2)\rho^{\nu+l} \Bigg\}. \end{aligned}$$

Collecting equal powers ρ with

$$u'' + \epsilon u = e^{-\lambda\rho}\left\{\sum_{\nu=-1}^{n_r} \rho^{\nu+l}\left[-2\lambda c_\nu(\nu+l+1)\right.\right.$$
$$\left.\left. + c_{\nu+1}(\nu+l+1)(\nu+l+2)\right]\right\}. \tag{17.28}$$

But (17.25) implies that $u'' + \epsilon u = [-\frac{2}{\rho} + \frac{l(l+1)}{\rho^2}]u(\rho)$; inserting (17.26) and expanding, the R.H.S. reads as

$$\frac{-2}{\rho}u + \frac{l(l+1)}{\rho^2}u = e^{-\lambda\rho}\left[-2\sum_{\nu=0}^{n_r} c_\nu\rho^{\nu+l} + l(l+1)\sum_{\nu=0}^{n_r} c_\nu\rho^{\nu+l-1}\right].$$

Now, shifting the sum again (with $c_{-1} = 0$), one finds

$$e^{-\lambda\rho}\left[-2\sum_{\nu=0}^{n_r} c_\nu\rho^{\nu+l} + l(l+1)\sum_{\nu=-1}^{n_r-1} c_{\nu+1}\rho^{\nu+l}\right]$$
$$= e^{-\lambda\rho}\sum_{\nu=-1}^{n_r} \rho^{\nu+l}\left[-2c_\nu + l(l+1)c_{\nu+1}\right].$$

Equating with (17.28),

$$-2\lambda c_\nu(\nu+l+1) + c_{\nu+1}(\nu+l+1)(\nu+l+2) = -2c_\nu + l(l+1)c_{\nu+1},$$

and, collecting terms,

$$c_{\nu+1}[(\nu+l+1)(\nu+l+2) - l(l+1)] + 2c_\nu[1 - \lambda(\nu+l+1)] = 0.$$

We need to simplify the coefficient of $c_{\nu+1}$. Set $a = \nu + 1$; we are left with

$$(\nu+l+1)(\nu+l+2) - l(l+1) = (a+l)(a+l+1) - l(l+1)$$
$$= a^2 + a(l+1) + al + l(l+1) - l(l+1) = a(a+2l+1) = (\nu+1)(\nu+2+2l).$$

Now solve for $c_{\nu+1}$:

$$c_{\nu+1} = 2\frac{\lambda(\nu+l+1) - 1}{(\nu+1)(\nu+2l+2)}c_\nu. \tag{17.29}$$

This recurrence relation like (16.10), for $\nu \to \infty$, gives an exponential series $f^{(n_r,l)}(\rho) \sim e^{2\lambda\rho}$. The series must terminate, otherwise the exponential growth would take us outside the space of square integrable functions. So, we must find $c_{n_r+1} = 0$ for $\nu = n_r$, where the *radial quantum number* n_r is the degree of the polynomial. The condition is

$$\lambda n = 1, \tag{17.30}$$

where

$$n = n_r + l + 1 \tag{17.31}$$

defines the so-called *principal quantum number* n. The hydrogenic wave function is labeled by n and the angular quantum numbers:

$$\psi_{n,l,m}(r,\theta,\phi) = R_{n,l}(r)Y_{l,m}(\theta,\phi).$$

The energy eigenvalues are $\epsilon = -\frac{1}{n^2}$, or, in view of (17.24),

$$E_n = -\frac{Z^2e^2}{2a_B n^2} = -\frac{13.59Z^2}{n^2}eV. \tag{17.32}$$

The energies depend exclusively on n; therefore, $2s$ and $2p$ are degenerate, $3s$, $3p$, $3d$ are degenerate and so on. The low entries of the energy level scheme are $1s$, $2s$, $2p$, $3s$, $3p$, $3d$, ... as follows.

Shell	n	Level	Angular momentum l	Allowed m
K	1	1s	0	$m = 0$
L	2	$2s$	0	$m = 0$
		$2p$	1	$m = -1, 0, 1$
M	3	$3s$	0	$m = 0$
		$3p$	1	$m = -1, 0, 1$
		$3d$	2	$m = -2, -1, 0, 1, 2$
N	4	$4s$	0	$m = 0$
		$4p$	1	$m = -1, 0, 1$
		$4d$	2	$m = -2, -1, 0, 1, 2$
		$4f$	3	$m = -3, -2, -1, 0, 1, 2, 3$
...	...	...	...	...

These levels agree with the experiment, apart from corrections that are small for Hydrogen. Relativistic corrections (spin–orbit interaction, dependence of mass from velocity, etc.) have a relative size $Z\alpha^2$, where Z is the atomic number and α is the fine structure constant. Other corrections arise from the electron spin magnetic moment, nuclear spin and quadrupole moment, and Quantum Electrodynamics. Spectroscopically, the photon frequencies that are emitted and absorbed are

$$\nu_{mn} = R\left(\frac{1}{n^2} - \frac{1}{m^2}\right),$$

where R is the Rydberg constant. Emission occurs in the decay $m \to n$, while the opposite transition leads to absorption. The series of lines with $n = 2$ was discovered by Balmer in 1885, and starts in the visible spectrum, with the line H_α with ν_{23} in the red, the blue line H_β with ν_{24} and the violet H_γ with ν_{25}; the series continues in the ultraviolet; later, the ultraviolet Lyman series $n = 1$; then the infrared series with $m = 3, 4, 5$; however, the explanation remained a mystery before the advent of Quantum Mechanics.[6] The explanation is

[6] An ad hoc model by N. Bohr in 1913 gave the same levels but nothing else.

$$\hbar\omega_{mn} = E_m - E_n = 13,59\text{eV}\left(\frac{1}{n^2} - \frac{1}{m^2}\right).$$

Eliminating λ in (17.29) by the condition $\lambda n = 1$, one finds the recurrence relation with the coefficients of $f^{(n_r,l)}$:

$$c_{\nu+1} = \frac{2}{n}\frac{\nu + l + 1 - n}{(\nu+1)(\nu+2l+2)}c_\nu. \tag{17.33}$$

This may be simplified by removing the factor $\frac{2}{n}$. To this end, we change scale. If a function $L(x) = \sum a_\nu x^\nu$ is defined by a recurrence relation $\frac{a_{\nu+1}}{a_\nu} = \xi(\nu)$, we can change scale by setting $L(x) \to L(sx)$, where s is arbitrary. Then,

$$L(sx) = \sum a'_\nu x^\nu, \quad a'_\nu = a_\nu s^\nu \Longrightarrow \frac{a'_{\nu+1}}{a'_\nu} = s\ \xi(\nu).$$

Therefore, we set $f^{(n_r,l)}(\rho) = L(\frac{2\rho}{n})$, and $L^{2l+1}_{n_r}$ has a simplified recurrence relation[7]

$$a_{\nu+1} = \frac{\nu + l + 1 - n}{(\nu+1)(\nu+2l+2)}a_\nu. \tag{17.34}$$

Now, from (11.12) one obtains

$$\begin{aligned} u(\rho) \quad &= e^{-\lambda\rho}\rho^{l+1}L^{2l+1}_{n_r}\left(\tfrac{2\rho}{n}\right);\ \lambda = 1/n, \\ R_{nl}(\rho) &= e^{-\frac{\rho}{n}}\rho^l L^{2l+1}_{n_r}\left(\tfrac{2\rho}{n}\right). \end{aligned} \tag{17.35}$$

Radial Functions

We can derive the radial functions from (17.34) by recalling that $\nu \le n_r$ and that $n = n_r + l + 1$. For $n = 1$, $n_r = l = 0$ e $\nu = 0$. Putting $a_0 = 1$ one gets $a_1 = 0$.

For $n = 2$, there are two cases: $n_r = 1$, $l = 0$, $n_r = 0$, $l = 1$.

For $n_r = 1, l = 0$, (17.34) becomes $a_{\nu+1} = \frac{\nu-1}{(\nu+1)(\nu+2)}a_\nu$. For $\nu = 0$, $a_1 = -a_0/2$; for $\nu = 1$, $a_2 = 0$. Therefore,

$$L^1_1(\rho) = 1 - \frac{1}{2}\rho,$$

[7] The *associated Laguerre polynomials* has the Rodrigues representation

$$L^k_n(x) = \sum_{m=0}^{n}(-1)^m\frac{(n-k)!}{(n-m)!(k+m)!m!}x^m.$$

.

and, up to a normalization constant,

$$R_{20}(\rho) = e^{-\frac{\rho}{2}}(1 - \frac{1}{2}\rho).$$

For $n_r = 0$, $l = 1$, there is only $\nu = 0$; Eq. (17.34) $a_{\nu+1} = \frac{\nu}{(\nu+1)(\nu+4)}a_\nu$ yields $a_1 = 0, l = 1$ and so

$$R_{21}(\rho) = e^{-\frac{\rho}{2}}\rho.$$

Here, I report few radial functions normalized with

$$\int_0^\infty |R_{n,l}(r)|^2 r^2 dr = 1.$$

$$R_{1,0} = \frac{2}{\sqrt{a_0^3}} e^{-\rho}$$
$$R_{2,0} = \frac{1}{\sqrt{2a_0^3}}(1 - \frac{\rho}{2})e^{-\frac{\rho}{2}}$$
$$R_{2,1} = \frac{1}{\sqrt{24a_0^3}}\rho e^{-\frac{\rho}{2}}$$
$$R_{3,0} = \frac{2}{\sqrt{27a_0^3}}(1 - \frac{2\rho}{3} + \frac{2\rho^2}{27})e^{-\frac{\rho}{3}}$$
$$R_{3,1} = \frac{8}{27\sqrt{6a_0^3}}(1 - \frac{\rho}{6})\rho e^{-\frac{\rho}{3}}$$
$$R_{3,2} = \frac{4}{81\sqrt{30a_0^3}}\rho^2 e^{-\frac{\rho}{3}}.$$

In general, $R_{nl}(r)$ has $n - 1$ nodes.

17.5.1 Coulomb Wave Functions

The continuum (positive energy) solutions called Coulomb waves are of interest in electron-nucleus scattering, and also in the description of the final state of photoelectrons (i.e., electrons ejected by an atom that absorbs a photon) and Auger electrons, i.e., electrons emitted by an atom while another electron jumps to a lower empty level. There is no special interest here to go into the solution of the radial equation in terms of confluent hypergeometric functions. However the interesting thing is that for large ρ, the wave function does not approach a free electron function, but gets a phase that depends on distance at all distances. The reason is that the Coulomb potential is long range.

17.5.2 Atomic Shells

The states with $n = 1, 2, 3, 4, \ldots$ form the so-called *atomic shells* $K, L, M, N, O, P \ldots$ which are observed *grosso modo* in all atoms. Actually, the description of many-electron atoms is much more involved, but to some extent one can understand the observed charge densities and binding energies in terms of atomic electrons moving in an effective field, due to the field of the nucleus plus the mean field of the others. This is the idea behind the Hartree–Fock method; see Chap. 21. Since $n_r \geq 0$, $l \in (0, n-1)$. Shell K has one orbital, shell L has 4 and shell M 9. How many in shell n? Since $\sum_{n=1}^{p} = \frac{p(p+1)}{2}$, the number of orbitals is $\sum_{l=0}^{n-1}(2l+1) = n^2$.

17.6 Lamb Shift and Radiative Corrections

In Schrödinger's theory, as we have seen, the 2s and 2p levels of H are degenerate, i.e., have the same energy. In the relativistic Dirac's theory, the Hydrogen states are denoted by nLJ and those with the same principal quantum number n and the same total angular momentum J (which includes the electron spin) remain degenerate even if they have different orbital angular momentum L. In particular, ${}^2S_{1/2}$ and ${}^2P_{1/2}$ levels remain degenerate.

In 1947, however, Lamb and Rutherford discovered the Lamb shift of atomic ${}^2S_{1/2}$ and ${}^2P_{1/2}$ levels. After producing atomic H, the levels n=2 were populated by scattering with electrons having 10.2 eV kinetic energy, but while the 2p states decay in about a nanosecond, the 2s level is metastable and lasts 1/7s. The excited atoms can be monitored and readily de-excited using radiofrequency corresponding to the $2s_{1/2} - 2p_{1/2}$ or $2s_{1/2} - 2p_{3/2}$ separation.

The measurement required great precision and involves a splitting of only about 1000 MHz.

This shift can be calculated in Quantum Electrodynamics (QED) with similar great precision. It is due to the fact that the atomic electrons are acted upon by the vacuum field, in addition to the nuclear electric field, and the vacuum field couples differently with the two states. This is a manifestation of the same zero point fields that produce the Casimir effect (Chap. 16). A simple description in terms of the quantum vacuum fields was proposed by Welton in 1948. Consider a box with side L. The energy density is

$$\rho(\omega) = \frac{1}{8\pi L^3}\int d^3\mathbf{r}[\mathbf{E^2} + \mathbf{B^2}] = \frac{\mathbf{1}}{\mathbf{4\pi L^3}}\int d^3\mathbf{r}[\mathbf{E^2}] = \frac{\mathbf{1}}{\mathbf{4\pi}}\langle \mathbf{E^2}\rangle. \tag{17.36}$$

On the other hand,

$$\int \rho(\omega)d\omega = \frac{1}{L^3}\sum_{\lambda,\mathbf{k}}\frac{\hbar\omega(\mathbf{k})}{\mathbf{2}}, \tag{17.37}$$

where the sum over polarizations brings a factor 2, and

$$\sum_{\mathbf{k}} \to \frac{L^3}{(2\pi^3)} \int d^3\mathbf{k}. \tag{17.38}$$

One obtains since $\omega = ck$

$$\frac{L^3}{2\pi^2 c^3} \int \omega^2 d\omega. \tag{17.39}$$

Thus the electromagnetic energy density in $d\omega$ is

$$\rho(\omega) = \frac{1}{2\pi^2 c^3 \omega^3}. \tag{17.40}$$

This argument has an upper limit $\omega < \omega_{max} = \frac{mc^2}{\hbar}$ because at higher frequencies the electromagnetic field produces electron–positron pairs, while here we are talking about a much lower frequency phenomenon. On the other hand, there is also a natural low frequency cutoff $\omega_{min} = \frac{c}{a_0}$ corresponding to long wavelengths.

Comparing with Eq. (17.36), we conclude that the vacuum electric fields producing fluctuations of the electron position are

$$\langle \mathbf{E^2}(\omega) \rangle = \frac{\mathbf{2}\omega^3}{\pi \mathbf{c^3}}. \tag{17.41}$$

This field produces fluctuations in the electron coordinates $\mathbf{r}$ and its intensity grows with frequency.

Classically, $\frac{d^2\mathbf{r}}{\mathbf{dt^2}} = \frac{e\mathbf{E}}{\mathbf{m}}$, which implies

$$\mathbf{r} = \frac{-\mathbf{eE}(\omega)}{\mathbf{m}\omega^2}; \tag{17.42}$$

the average shift vanishes, but the mean square shift due to fluctuations at frequency ω is

$$\langle \delta \mathbf{r^2}(\omega) \rangle = \left(\frac{\mathbf{e}}{\mathbf{m}\omega^2} \right)^2 \langle \mathbf{E^2}(\omega) \rangle \tag{17.43}$$

and the total is

$$\langle \delta \mathbf{r^2} \rangle = \int_{\omega_{min}}^{\omega_{max}} \langle \delta \mathbf{r^2}(\omega) \rangle \tag{17.44}$$

which is indeed about 1000 MHz, in agreement with experiment. Quantum Electrodynamics accurately predicts the upwards shift by 1040 MHz of 2s and 8 MHz of $2p_{3/2}$ and the downward shift of $2p_{1/2}$ by 17 MHz.

17.7 Periodic Potentials, Floquet's Theorem and Bloch's Functions: Tight Binding Bands of Solids

The simplest quantum model for a solid assumes independent particles; the electron sees a potential $U(\vec{r})$ with the lattice periodicity. The direct lattice (or Bravais lattice) is a set of points characterized by translations $\vec{a}_1$, $\vec{a}_2$, $\vec{a}_3$. Each lattice point belongs to a so-called Wigner–Seitz cell, such that using the full set of translation vectors we can take it to any other Wigner–Seitz cell, neatly filling all the space without overlapping.

Naturally enough one tries to start by neglecting all the electron–electron interactions, which appears a gross oversimplification. In the early days of Solid State Physics, this was justified by the lack of means to do more realistic calculations, and it was discovered that the approach had some success in materials like Si and Al while it failed completely in other cases. The present understanding is better explained in Sect. 19.4, but at present we are interested to see where this optimistic approach can be carried on. So, we drop the interactions and the relativistic complications.

The electronic Hamiltonian is of the form

$$H = \frac{\hat{p}^2}{2m} + U(\vec{r}). \tag{17.45}$$

For any translation vector $\vec{R}$ connecting two equivalent points, one assumes that $U(\vec{r} + \vec{R}) = U(\vec{r})$. This is a periodic potential with the features of the potential marking the lattice sites. The solutions of the Schrödinger equations for $U \to 0$ may be taken to be plane waves $\psi_k(\vec{r}) \propto e^{i\vec{k}\cdot\vec{r}}$, while in the periodic U we may ask that the probability density be periodic. A particle in the crystal has a reduced symmetry compared to a free particle. It turns out that the eigenfunctions of the Hamiltonian can be chosen to have the same periodicity as the crystal and in the Fourier expansion of an eigenfunction only plane waves having full crystal periodicity can enter. This means that such plane waves must satisfy the condition

$$e^{i\vec{k}\cdot\vec{R}} = 1, \tag{17.46}$$

and such vectors form the reciprocal lattice. The Wigner–Seitz cell of the reciprocal lattice is the first Brillouin zone. The primitive vectors $\vec{b}_i$ of the reciprocal lattice must satisfy

$$\vec{b}_i \cdot \vec{a}_j = \delta_{ij}. \tag{17.47}$$

Floquet's theorem states that the wave functions may be taken of the form

$$\psi_{n\vec{k}}(\vec{r}) = e^{i\vec{k}\cdot\vec{r}} u_{n\vec{k}}, \tag{17.48}$$

where n is a band index and $|u_{n\vec{k}}|^2$ is lattice-periodic. I omit the proof, which is readily available.

Typically one finds allowed and forbidden energy bands. In this independent electron approximation, one can model the band structure of solids using various approximate schemes that allow to compute densities of states, band gaps and other observables. One simple such scheme is known as the tight binding approximation. For example, one can consider a chain of $N \to \infty$ atoms assuming that the diagonal matrix element on each site is $\langle n|H|n\rangle = E_0$, $\langle n-1|H|n\rangle = \langle n+1|H|n\rangle = -\Delta$, and the overlap between neighbor sites is S. The ansatz for the wave function is in line with the Floquet theorem:

$$|k\rangle = \frac{1}{\sqrt{N}} \sum_n e^{inka} |n\rangle, \tag{17.49}$$

where a is the lattice parameter, that is, the nearest neighbor distance. One easily finds the band structure

$$E(k) = \frac{E_0 - 2\Delta\cos(ka))}{1 + 2\cos(ka)}.$$

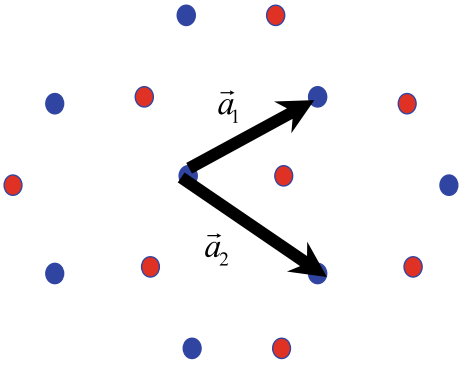
17.7.1 Band Structure of Graphene

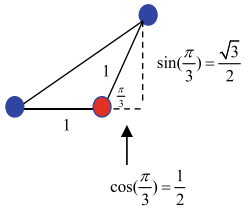
Andre Geim and Konstantin Novoselov received the Nobel prize in 2010 for discovering a preparation method and the unique mechanical and electrical properties (including superconductivity) of his two-dimensional material. Graphene is a two-dimensional hexagonal lattice which is bipartite (the sites can be represented by red and blue dots, such that all the neighbors of a blue dot are red and all the neighbors of a red dot are blue) (Fig. 17.1). From each blue (red) dot, one can hop to a red (blue) dot by the following vectors:

$$\vec{\delta}(1) = a(1,0);\ \vec{\delta}(2) = a(-\frac{1}{2}, \frac{\sqrt{3}}{2});\ \vec{\delta}(3) = a(-\frac{1}{2}, -\frac{\sqrt{3}}{2})).$$

Note that this choice of the geometry privileges the x axis. Let $a \sim 0.142\,\text{nm}$ denote the side of the hexagon (connecting a blue and a red dot). The primitive lattice vectors (connecting dots of the same color) may be written in Cartesian coordinates $a_1 = a(\frac{3}{2}, \frac{\sqrt{3}}{2}), a_2 = a(\frac{3}{2}, \frac{-\sqrt{3}}{2})$. By definition, the reciprocal lattice vectors $\vec{G}$ must be such that $\vec{G}.\vec{a}_1 = 0, \pm 2\pi, \pm 4\pi, \cdots$, and the same for $\vec{a}_2$. We may pick $\vec{G}_1 = \frac{4\pi}{3a}(\frac{1}{2}, \frac{\sqrt{3}}{2})$, $\vec{G}_2 = \frac{4\pi}{3a}(\frac{-1}{2}, \frac{\sqrt{3}}{2})$.

There are 6 smallest equivalent lattice vectors, and one can take $\vec{G}_3 = -\vec{G}_2, \vec{G}_4 = -\vec{G}_1, \vec{G}_5 = \frac{4\pi}{3a}(1,0)$ and $\vec{G}_6 = -\vec{G}_5$. To obtain the Brillouin Zone, draw the smallest G vectors and the straight lines through the centers: the inte-





$\Rightarrow$ Cartesian components are :

$$\vec{a}_1 = a(\frac{3}{2}, \frac{\sqrt{3}}{2}),\ \vec{a}_2 = a(\frac{3}{2}, -\frac{\sqrt{3}}{2}),$$

and so $|\vec{a}_1| = |\vec{a}_2| = \sqrt{3}a$

Fig. 17.1 Top: the Graphene layer is a bipartite lattice (red and blue atoms) and the unit vectors a_1 and a_2 connect atoms of the same sublattice. Bottom: Cartesian components of the unit vectors

rior is the Brillouin Zone, and the center of it is called the Γ point. The reciprocal lattice is also hexagonal.

Setting the origin of energies at the carbon atom energy level, the tight binding Hamiltonian is (understanding the spin index, which is diagonal)

$$\hat{H} = J_0 \sum_{\mathbf{i}}^{3} \sum_{\mathbf{j}} (|a_{\mathbf{j}} + \delta(\mathbf{i})\rangle\langle b_{\mathbf{j}}| + h.c. \tag{17.50}$$

where $\sum_{\mathbf{j}}$ sums over the lattice sites, h.c. stands for the Hermitian conjugate, bold face denotes vectors and $J_0 \sim 3$ eV. A Fourier transform solves the problem. Inserting $|a_{\mathbf{j}}\rangle = \sum_{\mathbf{k}} |a_{\mathbf{k}}\rangle e^{-i\mathbf{k}\cdot\mathbf{j}}$, and using

$$\gamma(\mathbf{k}) = \sum_{\mathbf{n=1}}^{\mathbf{3}} \mathbf{e}^{\mathbf{i}\mathbf{k}\cdot\delta(\mathbf{n})} = \mathbf{e}^{\mathbf{i}\mathbf{k}_\mathbf{x}} + \mathbf{2}\mathbf{e}^{-\mathbf{i}\frac{\mathbf{k}_\mathbf{x}}{\mathbf{2}}} \cos\left(\frac{\sqrt{\mathbf{3}}\mathbf{k}_\mathbf{y}}{\mathbf{2}}\right), \tag{17.51}$$

we find

$$H = \int d^2k (a^*(\mathbf{k}) \mathbf{b}^*(\mathbf{k})) \begin{bmatrix} 0 & J_0\gamma(\mathbf{k}) \\ J_0\gamma(\mathbf{k}) & 0 \end{bmatrix} \begin{pmatrix} a(\mathbf{k}) \\ b(\mathbf{k}) \end{pmatrix}. \tag{17.52}$$

The band structure is given by

$$E(\mathbf{k}) = \pm \mathbf{J_0} \sqrt{\mathbf{1} + \mathbf{4} \cos\left(\frac{\mathbf{3}}{\mathbf{2}} \mathbf{a} \mathbf{k_x}\right) \cos\left(\frac{\sqrt{\mathbf{3}}}{\mathbf{2}} \mathbf{a} \mathbf{k_y}\right) + \mathbf{4} \cos\left(\frac{\sqrt{\mathbf{3}}}{\mathbf{2}} \mathbf{a} \mathbf{k_y}\right)^{\mathbf{2}}} \tag{17.53}$$

where the + sign refers to the empty upper conduction band and the - sign to the filled lower band (the valence band). $\gamma(\mathbf{k})$ vanishes at the reciprocal lattice vectors, so there is no gap between the valence band and the conduction band. The absence of a gap between valence and conduction bands is typical of metals. The dependence of energy on wave vector near those point is linear, mimicking the Dirac equation. Na Xin and colleagues, in a paper published on Nature Vol. 616, page 270, reported the unusual behavior of these "Dirac electrons" in strong magnetic field, producing giant magnetoresistance.

17.7.2 Su–Schrieffer–Heeger Model

A remarkable paper by Su, W. P.; Schrieffer, J. R.; Heeger, A. J. entitled "Solitons in Polyacetylene" appeared in Physical Review Letters in 1979. Apparently, it is a simple one-dimensional model of a chain with alternating strong and weak bonds, which could be suitable for Polyacetylene. However this was a seminal paper, because the model has several nontrivial aspects, anticipating the current interest in topological insulators. The Hamiltonian reads

$$H = v \sum_{n=1}^{N} |n, B\rangle\langle n, A| + w \sum_{n=1}^{N} |n+1, A\rangle\langle n, B| + h.c. \tag{17.54}$$

where h.c. stands for the Hermitian conjugate. For $N \to \infty$, a Fourier transform

$$|n, A\rangle = \frac{1}{\sqrt{N}} \sum_k e^{ikn} |k, A\rangle, |n, B\rangle = \frac{1}{\sqrt{N}} \sum_k e^{ikn} |k, B\rangle \tag{17.55}$$

yields

$$H(k) = \begin{pmatrix} 0 & v + e^{ik} w \\ v + e^{-ik} w & 0 \end{pmatrix}; \tag{17.56}$$

hence, the energy eigenvalues are

$$\epsilon = \pm\sqrt{v^2 + w^2 + 2vw\cos(k)} \tag{17.57}$$

with the eigenstates

$$|k\pm\rangle = \frac{\pm e^{i\phi_k}|k, A\rangle + |k, B\rangle}{\sqrt{2}}, \tag{17.58}$$

where

$$\tan(\phi_k) = \frac{w\sin k}{v + w\cos k}. \tag{17.59}$$

Other interesting properties of this model will be presented in Chaps. 23 and 26. For the moment, we can observe that if $v = w$, the model reduces to a simple metallic chain (no gap) while otherwise the model shows two bands separated by a gap; so, the model might represent an insulator with a valence band and a conduction band separated by an energy gap.

Reference

Pauli, W.: Z. Phys. **36**, 336 (1926)

Chapter 18
Spin and Magnetic Field

An electron has radius 0 (as far as we know), a quantized angular momentum (spin) $\frac{1}{2}\hbar$ and a magnetic moment. Let us discover how.

18.1 Magnetic Moment and Angular Momentum in Classical Physics

Classically, a point charge that circulates on a ring of radius r produces a current $i = \frac{ev}{2\pi r}$, which causes a magnetic dipole moment normal to the ring, $\vec{\mu} = \frac{i}{c} S \vec{n}$, $S = \pi r^2$, in obvious notation;

$$\vec{\mu} = \frac{ev}{2c} r \vec{n} = \frac{e}{2c} \vec{r} \wedge \vec{v} = \frac{e}{2\,mc} \vec{L}\,, \tag{18.1}$$

where $\vec{L}$ is the angular momentum. The so-called gyromagnetic factor $\frac{e}{2mc}$ changes an angular momentum to a magnetic moment. The two are always parallel, in all theories. The magnetic dipole produces a magnetic field. We can take for the vector potential the expression

$$\vec{A} = \frac{\vec{\mu} \wedge \vec{r}}{r^3}. \tag{18.2}$$

In a uniform external magnetic field $\vec{B}$, the classical dipole has energy

$$E = -\vec{\mu} \cdot \vec{B}\,. \tag{18.3}$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_18

The angle θ between the magnetic moment and $\boldsymbol{B}$ is a constant of the motion (since it takes energy to modify θ).

A dipole in a space-dependent field feels a force

$$\overrightarrow{F} = -\overrightarrow{\nabla} E = \overrightarrow{\nabla}[\overrightarrow{\mu} \cdot \overrightarrow{B}]. \tag{18.4}$$

Measurement of Microscopic Magnetic Moments

Suppose we have a beam of unknown magnetic moments μ, with no charge, and known mass m. One can measure μ by letting the beam cross an inhomogeneous field in a short time Δt and measuring the deflection. (The electric charge would be deflected by the Lorentz force, which is much larger than the force on the dipole, and this is the reason for excluding it.)

If l is the width of the inhomogeneous field region where $\frac{\partial B}{\partial z} \neq 0$ and the magnetic moment has a known velocity v, it spends a time $\Delta t = \frac{l}{v}$ in the field. Then,

$$m\Delta v \approx F\Delta t, \text{ quindi } \Delta v \approx \mu_z \frac{\partial B_z}{\partial z} \frac{l}{mv},$$

and the deflection must be

$$\Delta\theta \approx \frac{\Delta v}{v} \approx \mu_z \frac{\partial B_z}{\partial z} \frac{l}{mv^2}. \tag{18.5}$$

18.2 The Stern–Gerlach Experiment

In 1922, Stern–Gerlach conducted one of the most important and astonishing experiments in history. They produced an atomic beam of Ag and sent it through an inhomogeneous magnetic field. Their question was: do the electronic currents in the atom produce a magnetic moment? Ag had been chosen because it was monovalent, so it was possible that the only chemically active electron circulating in a larger orbit played a role in the experiment.[1] The beam was obtained from metal vapor produced in a crucible; the whole apparatus operated in vacuo and Ag was known to remain atomic. The beam was collimated and, in the absence of a field, formed a spot on a photographic plate. In Fig. 18.1, a schematic drawing of the geometry is shown. The shape of the magnet was optimized in order to produce an inhomogeneous field, such that the trajectory depended on the orientation of the atomic dipole.

The plan of the experiment was to achieve a measurement of the dipole μ: it was expected that the beam should broaden into a cone, due to the random orientation of the dipoles in the crucible and in the beam; from the angle θ at the vertex of the cone one could deduce μ according to Eq. (18.5). Stern and Gerlach had planned

[1] The Bohr model predicted orbits, and classically there is no other way than a current to produce a magnetic moment; but Nature had a surprise in store.

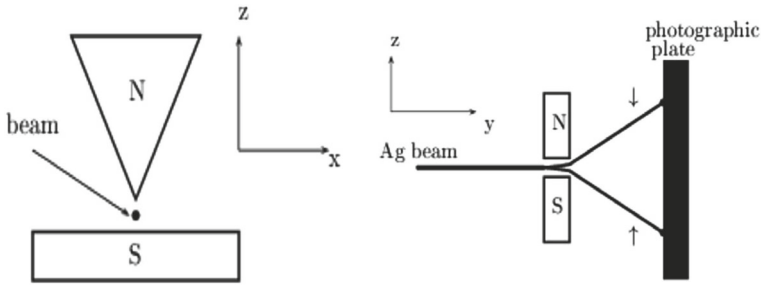


Fig. 18.1 Schematic sketch of the Stern–Gerlach apparatus. Left: seen from the crucible, along the y axis; the magnetic field is strongly inhomogeneous due to the pointed shape of the North pole. Right: seen from the direction of the x axis

their experiment very well, from the classical point of view, and the atoms did show a moment of the expected order of magnitude. But the beam does not broaden; it splits into two, equally intense branches, one with the moment parallel to the field and the other with opposite moment.

Each branch consists of atoms with the same moment orientation, and the same μ_z. If one of these beams is sent to a second apparatus oriented like the previous one, it can be deflected, but does not split again. This is expected, since all the atoms are in the same state, see the same field, and there is no reason for a splitting. But, if the second apparatus is oriented differently, instead of measuring a decreased component due to the angle between the old and new direction, a new splitting occurs. Why? This fact cannot be understood in terms of classical magnetic moments. There is, however, an optical analogue (see the next subsection).

It was clear that the magnetic moment of Ag is due to the valence electron and that it is associated with an angular momentum $\overrightarrow{J}$. I am not using the orbital angular momentum $\overrightarrow{L}$ because it is not enough. I am not using the symbol $\overrightarrow{L}$ of the orbital angular momentum because it is not appropriate, here; no L value gives two components. In fact, it is due to the electronic spin,[2] the one of the chemically active optical electron. Summarizing the results of the experiment (and using also the results of Chap. 17):

1. The Ag atom has a magnetic moment μ and the values of μ_z along any direction are quantized.
2. Since $\overrightarrow{\mu} \propto \overrightarrow{J}$, the values of the components J_z in the field direction are quantized, too.
3. There are $2j + 1$ possible values of m; the experiment with Ag shows that the values are 2, so according to rules of Chap. 17 the angular momentum is $\frac{1}{2}$; however, no spherical harmonic has a fractional angular momentum. The relativistic

[2] Most nuclei have spins and magnetic moments; the nuclear magnetic moments, however, are negligible compared to the electronic ones. This is related to the proton and neutron masses, which are much heavier than the electron.

Dirac's theory naturally explains the spin and the associated magnetic moment, but the present theory accepts them as facts and W: Pauli provided a coherent formalism for the non-relativistic approximation (see below).

4. The magnetic moment is found either up or down in any direction one measures it.
5. The outcome of the experiment cannot be understood in terms of classical point masses carrying a magnetic moment. Indeed, the direction of the magnetic moments is not a property of the Ag atoms, but is determined by the field.

This shows some essential features of Quantum Mechanics. The interpretation is as follows:

1. Every angular momentum due to some rotation in space (orbital angular momentum) is associated with $\overrightarrow{L}$ and has integer eigenvalues; besides, there is spin, which has no classical counterpart. In general, we shall denote angular momenta of any origin by $\overrightarrow{J}$ and the component by m_j.
2. The Ag atoms do not have a sharply defined component μ_z of the magnetic moment when they form the beam, but only when one measures μ_z.
3. The act of measurement produces a collapse of the wave function and two branches with a well-defined component, μ_z, say; μ_x and μ_y are not defined.
4. Any measurement along a different direction produces a new collapse of the wave function (see Fig. 18.2).

If one knows that an electron is in some point $\overrightarrow{x}$, the information is not complete. It has two orthogonal ways to be there. The two ways correspond to the two components of $S_z = \pm\frac{\hbar}{2}$, where z is the direction privileged by the experiment. No motion of particles can realize a fractional angular momentum. No spherical harmonic is relevant to it. Actually, the spin is an intrinsic angular momentum, intrinsic like mass or charge, with a z component that can be exchanged with the outside world in fixed amounts, while z is in a direction that can be chosen at will.

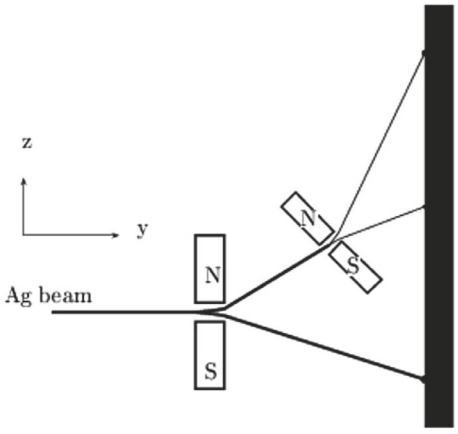


Fig. 18.2 A second Stern–Gerlach apparatus with the field in a new direction splits the beam again

18.2.1 Spin of Elementary Particles and Bound States

Nobody had imagined the electron spin before the discovery, but there is an analogy with the polarization of electromagnetic waves which was known from Maxwell's equations. An electromagnetic wave carries momentum and therefore also orbital angular momentum. The orbital angular momentum is

$$\boldsymbol{L} = \frac{1}{4\pi} \int d^3 r \boldsymbol{r} \wedge [\boldsymbol{E} \wedge \boldsymbol{B}], \tag{18.6}$$

in Gaussian units. Besides this orbital angular momentum, the wave also carries spin angular momentum if it is left or right circularly polarized, like for instance a wave propagating long z with $E_x = E_0 \cos(\omega t)$, $E_y = \pm E_0 \sin(\omega t)$. Even in this case no object turns, and the angular momentum is intrinsic. We observe the spin through the angular momentum that photons can exchange with matter. A left polarized photon carries an angular momentum $+\hbar$. Beth[3] reports a measurement of the spin using a special film that converts a right polarized wave to left polarized. The film was suspended to a quartz fiber and exposed to left polarized radiation, and the angular momentum was mechanically measured. Spin $S = 1$ implies three states with M $= -1, 0, 1$, but this rule holds for massive particles that have a rest reference and so can spin in any direction. The photon is different because it has no rest reference; so it can spin only in the direction of the wave vector and the only values are M $= -1, 1$. One can perform an optical experiment which bears some analogy with the Stern–Gerlach experiment using a Wollaston prism, invented by William Hyde Wollaston, that separates an unpolarized beam into two orthogonal linearly polarized outgoing beams. This is based on the property of birefringence of some anisotropic crystals like calcite, quarz, sapphire, ruby and $LiNbO_3$ whose refraction index depends on the polarization and on the direction of propagation.

Not only electrons and photons, but all fundamental particles have spin quantum numbers. The electron, the muon and the tau have spin 1/2, and together they are classified as leptons; each of them is associated with a spin 1/2 neutrino (electron neutrino, muon neutrino and tau neutrino) . These were long supposed to be massless like the photon, before the discovery that they change flavor in flight. There are six flavors of quarks (down, up, strange, charm, bottom and top), also having spin 1/2. The mediators have spin 1 (gluons for the strong interaction, W^+, W^- , Z^0 for the weak and of course photons for the electromagnetic interaction). Baryons and mesons are compound particles. Similarly, the spins of atomic nuclei are due to the spins and angular momenta of the costituent particles. The even–even nuclei (even numbers of protons and neutrons, like ^{12}C and ^{16}O) have no spin; the even–odd nuclei have spin 1/2 , like ^{1}H and ^{31}P ; the odd–odd nuclei like ^{2}H and ^{14}N have spin 1. There are however high spin nuclei like ^{123}Sb (spin 7/2).

In atomic physics, the interaction of the electrons with nuclear spins produces the hyperfine structure of the levels. Since there is a magnetic moment associated with

[3] Beth (1936).

every spin, one can obtain information about the geometry and electronic structure of molecules and solids by producing transitions between the levels of the system in an external magnetic field. The electron paramagnetic magnetic resonance (epr) and the nuclear magnetic resonance (nmr) have important applications in condensed matter physics and also medical applications.

18.3 Pauli Matrices

Now we must develop the formalism for spin 1/2. For the orbital angular momentum $\overrightarrow{L}$, and for integer spins, we know from Section (13.8) the matrix elements

$$\langle l_1 m_1 | L^2 | l_2 m_2 \rangle = \hbar^2 l_1 (l_1 + 1) \delta_{l_1, l_2} \delta_{m_1, m_2},$$

$$\langle l_1 m_1 | L_z | l_2 m_2 \rangle = \hbar m_1 \delta_{l_1, l_2} \delta_{m_1, m_2},$$

$$\langle l_1 m_1 | L_\pm | l_2 m_2 \rangle = \hbar \sqrt{l_1 (l_1 + 1) - m_2 (m_2 \pm 1)} \delta_{l_1, l_2} \delta_{m_1 \pm 1, m_2},$$

with integer l, m. None of these is suitable for half-integer spin; we need the spinor representations,[4] which correspond to half-integer angular momenta $(\frac{1}{2}, \frac{3}{2}, \frac{5}{2} \cdots)$. We shall deal with spin $\frac{1}{2}$ in detail.

For the electron spin degree of freedom, we need a formalism without the derivatives $(\frac{\partial}{\partial r}, \frac{\partial}{\partial \theta}, \frac{\partial}{\partial \phi})$, since nothing is rotating in space. We shall take the z axis as the quantization axis; in a Stern–Gerlach experiment it is naturally identified with direction of the magnetic field.

For a spin $\frac{1}{2}$, the dimension of Hilbert space is 2, since there are two orthogonal states $|\alpha\rangle = |S_z = \frac{1}{2}\rangle$ and $|\beta\rangle = |S_z = -\frac{1}{2}\rangle$.

These are called conventionally up and down spin and will be taken as the basis of the spin Hilbert space.[5] Letting $S_z = \frac{\hbar}{2}\sigma_z$, the basis states and the z component are chosen as

$$|\alpha\rangle \equiv |\frac{1}{2}\rangle \to \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \ |\beta\rangle \equiv |-\frac{1}{2}\rangle \to \begin{pmatrix} 0 \\ 1 \end{pmatrix}; \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{18.7}$$

The two-component wave functions, called *spinors*, are written like complex column vectors $\psi = \begin{pmatrix} a \\ b \end{pmatrix}$; the Hermitian conjugate is

[4] The following theory is due to Wolfgang Pauli, that we have met already; he was one of the most important theoreticians of the twentieth century.

[5] In general the system has many other degrees of freedom and the full Hilbert space is the tensor product of the spin Hilbert space times the Hilbert space arising from the other degrees of freedom.

$$\psi^\dagger = \begin{pmatrix} a \\ b \end{pmatrix}^\dagger = \begin{pmatrix} a^* & b^* \end{pmatrix}. \tag{18.8}$$

The normalization condition reads

$$\psi^\dagger \psi = |a|^2 + |b|^2 = 1. \tag{18.9}$$

Introducing shift operators

$$S_+ = \hbar \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad S_- = \hbar \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \tag{18.10}$$

one can switch spin:

$$S_+|\alpha\rangle = 0, \quad S_-|\alpha\rangle = \hbar|\beta\rangle, S_-|\beta\rangle = 0, \quad S_+|\beta\rangle = \hbar|\alpha\rangle. \tag{18.11}$$

By analogy with the orbital angular momentum, from the shift operators $S_\pm$, one gains the x and y components:

$$\begin{aligned} S_+ &= S_x + iS_y \\ S_- &= S_x - iS_y. \end{aligned} \tag{18.12}$$

Here is the spin representation in terms of Pauli matrices:

$$\overrightarrow{S} = \hbar \frac{1}{2} \overrightarrow{\sigma}, \tag{18.13}$$

with

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{18.14}$$

It is easy to check that the square of the Pauli matrices is the unit matrix and that they anticommute,

$$\left[\sigma_i, \sigma_j\right]_+ = 2\delta_{ij}, \tag{18.15}$$

where the indices run over the three components. The identification of $\overrightarrow{S}$ with angular momentum is legitimate, since

$$\overrightarrow{S} \wedge \overrightarrow{S} = i\hbar \overrightarrow{S}, \tag{18.16}$$

which holds, since

$$\left[\sigma_x, \sigma_y\right]_- = 2i\sigma_z, \tag{18.17}$$

and so on. Moreover,

$$S^2 = \frac{3}{4}\hbar^2 = \frac{1}{2}\left(\frac{1}{2}+1\right)\hbar^2. \tag{18.18}$$

An electron can go from α and β or back by adsorbing or creating a spin-1 photon in magnetic resonance experiments. It is a magnetic dipole transition. If the electron spin is localized over an atom, the wavelength of the photon is much larger; in such cases, the orbital angular momentum of the photon is practically zero and the photon spin changes by 1.

The Stern–Gerlach experiment shows that the spin component is $\pm\frac{1}{2}\hbar$ in any direction, and the analogy with the spin of a tennis ball is misleading. The Pauli formalism describes this state of affairs well.

The Hermitian matrix that represents the spin in the direction

$$\overrightarrow{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \tag{18.19}$$

is

$$\overrightarrow{S}\cdot\overrightarrow{n} = \frac{1}{2}\hbar\begin{pmatrix}\cos\theta & \sin\theta e^{-i\phi}\\ \sin\theta e^{i\phi} & -\cos\theta\end{pmatrix}. \tag{18.20}$$

The eigenvectors are

$$|\uparrow, \overrightarrow{n}\rangle = \begin{pmatrix}\cos\frac{\theta}{2}\\ \sin\frac{\theta}{2}e^{i\phi}\end{pmatrix}, \tag{18.21}$$

$$|\downarrow, \overrightarrow{n}\rangle = \begin{pmatrix}-\sin\frac{\theta}{2}e^{-i\phi}\\ \cos\frac{\theta}{2}\end{pmatrix}. \tag{18.22}$$

For example, the physical meaning of the upper component $\cos\frac{\theta}{2}$ of $|\uparrow, \overrightarrow{n}\rangle$ is the following. Suppose a beam of particles is prepared in the state $|\uparrow, \overrightarrow{n}\rangle$, by a Stern–Gerlach experiment with the inhomogeneous field along the $\overrightarrow{n}$ axis. Then, suppose the z component of the spin is measured, maybe using a new Stern–Gerlach apparatus with the field along z. Then, $\cos\frac{\theta}{2} = \langle\alpha|\uparrow, \overrightarrow{n}\rangle$ is the probability amplitude of up spin.

Problem 18.1 Verify the eigenvectors.

Solution 18.1

$$\begin{pmatrix}\cos\theta & \sin\theta e^{-i\phi}\\ \sin\theta e^{i\phi} & -\cos\theta\end{pmatrix}\begin{pmatrix}\cos\frac{\theta}{2}\\ \sin\frac{\theta}{2}e^{i\phi}\end{pmatrix} = \begin{pmatrix}\cos(\theta)\cos(\frac{\theta}{2}) + \sin(\theta)\sin(\frac{\theta}{2})\\ \sin(\theta)\cos(\frac{\theta}{2})e^{i\phi} - \cos(\theta)\sin(\frac{\theta}{2})e^{i\phi}\end{pmatrix}.$$

Substituting the identities

$$\cos(\theta) = \cos\left(\frac{\theta}{2}\right)^2 - \sin\left(\frac{\theta}{2}\right)^2,\ \sin(\theta) = 2\cos\left(\frac{\theta}{2}\right)\sin\left(\frac{\theta}{2}\right)$$

one finds

$$= \begin{pmatrix} \cos(\frac{\theta}{2}) \left[\cos(\frac{\theta}{2})^2 - \sin(\frac{\theta}{2})^2 + 2\sin(\frac{\theta}{2})^2\right] \\ \sin\frac{\theta}{2} e^{i\phi} \left[2\cos(\frac{\theta}{2})^2 - \cos(\frac{\theta}{2})^2 + \sin(\frac{\theta}{2})^2\right] \end{pmatrix} = \begin{pmatrix} \cos(\frac{\theta}{2}) \\ \sin(\frac{\theta}{2}) e^{i\phi} \end{pmatrix}.$$

The True Origin of Electron Spin

There is no premonition of spin in the postulates. The Pauli theory looks like a brilliant patch; in fact it is *phenomenological*. This is how one describes theories that properly explain some phenomena, but also show the need for a broader and more satisfactory framework. Actually in the Schrödinger theory one starts with the classically known observables and Pauli has shown that the novel concept of spin fits well in the theory. A much more satisfactory formulation arises from the relativistic formulation. For the electron, this is Dirac's theory.

18.3.1 Euler-Angle Representation

It is useful to know an alternative way of defining a direction in space, using the well-known Euler angles ϕ, θ, ψ. Letting $|\uparrow\rangle$ denote a spin up (typically, in the positive zeta direction of a Cartesian reference), one can write down the rotation of a spin $\frac{1}{2}$ in the direction designated by Euler angles as follows:

$$g(\phi, \theta, \psi) = e^{-i\phi S_3} e^{-i\phi S_2} e^{-i\psi S_3} |\uparrow\rangle. \tag{18.23}$$

Here, ϕ and ψ vary in $(0, 2\pi)$ while θ in $(0, \pi.)$ The well-known group of all these rotations is called SU(2).

18.3.2 How to Rotate a Spin 1/2?

How to rotate a spinor like, for instance $|\uparrow, \overrightarrow{n}\rangle = \begin{pmatrix} \cos\frac{\theta}{2} \\ \sin\frac{\theta}{2} e^{i\phi} \end{pmatrix}$ by an angle $\Delta\phi$ around z axis? Well, ϕ is the *longitude*, and one could quite reasonably argue simply that $\phi \to \phi + \Delta\phi$. This is *almost true,* but there is a subtle, striking complication. In Chap. 13, we determined how to achieve the rotation of an orbital acting not on the argument $\boldsymbol{x}$ of $\psi(\boldsymbol{x})$ but on ψ. This requires a unitary operator $R_\phi = e^{-i\frac{\Delta\phi \cdot L}{\hbar}}$. By analogy, the generator of the infinitesimal rotations of spin is $\overrightarrow{S} = \frac{1}{2}\hbar\overrightarrow{\sigma}$ and the operator that does the rotation is

$$R_\phi = e^{-i\Delta\phi \cdot \frac{S}{\hbar}}. \tag{18.24}$$

At the exponent, we find a 2×2 matrix, which is easy to diagonalize and exponentiate. The operator (18.24) that rotates around the z axis reads as

$$R^{(z)}_{\Delta\phi} = e^{-i\frac{\Delta\phi}{2}\sigma_z} = e^{-i\frac{\Delta\phi}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}}. \tag{18.25}$$

Diagonal matrices are exponentiated immediately:

$$\begin{pmatrix} \mu & 0 \\ 0 & \nu \end{pmatrix}^2 = \begin{pmatrix} \mu^2 & 0 \\ 0 & \nu^2 \end{pmatrix},$$

and so

$$e^{\begin{pmatrix} \mu & 0 \\ 0 & \nu \end{pmatrix} t} = \begin{pmatrix} e^{\mu t} & 0 \\ 0 & e^{\nu t} \end{pmatrix}, \tag{18.26}$$

and the rotation operator is

$$R^{(z)}_{\Delta\phi} = \begin{pmatrix} e^{-i\frac{\Delta\phi}{2}} & 0 \\ 0 & e^{i\frac{\Delta\phi}{2}} \end{pmatrix}. \tag{18.27}$$

Therefore,

$$R^{(z)}_{\Delta\phi}|\uparrow, \overrightarrow{n}\rangle = \begin{pmatrix} e^{-i\frac{\Delta\phi}{2}}\cos\frac{\theta}{2} \\ e^{i\frac{\Delta\phi}{2}}\sin\frac{\theta}{2}e^{i\phi} \end{pmatrix} = e^{-i\frac{\Delta\phi}{2}}\begin{pmatrix} \cos\frac{\theta}{2} \\ \sin\frac{\theta}{2}e^{i[\phi+\Delta\phi]} \end{pmatrix}. \tag{18.28}$$

Thus, $R^{(z)}_{\Delta\phi}$ increments ϕ, as was his duty, but also gives an overall phase to the spinor. The implication is that $R_{2\pi} = -1$: a rotation by 2π changes the sign of the spinor! This is a fundamental property of half-integer spin particles. The set of all the matrices (18.24) with the rule of matrix multiplication constitutes a mathematical structure called the SU(2) Group (S stands for special, and means that all the matrices have a unit determinant). The O(3) Group made with the matrices (13.21) does almost the same job, but on scalar wave functions instead. In both cases, all the rotations can be represented; however, in $SU(2)$, there are two different operators for each rotation. Traditionally, $SU(2)$ is called the covering Group of $O(3)$. The (-1) factor does not change the state, but if one could superpose two beams then the (-1) factor affecting one of them would lead to destructive interference. Such an experiment, however, has not yet been reported.

For a general rotation, one can choose the rotation axis as the z axis; otherwise one can build the rotations around the other coordinate axes. As in Chap. 13, we must take into account that rotations around different axes do not commute. For a rotation by an angle α around an arbitrary axis, one can write

$$R_\alpha = R_\phi^{(z)} R_\theta^{(x)} R_\psi^{(z)}, \tag{18.29}$$

where $\phi \in (0, 2\pi)$, $\theta \in (0, \pi)$ and $\psi \in (0, 2\pi)$ are Euler angles representation of the rotation.

18.3.3 Spin Magnetic Moment and Pauli Equation

The classical Eq. (18.1) associates a magnetic moment with the angular momentum. Since within the electron, there are no currents, we cannot compute the electron moment classically. In Dirac's Relativistic Quantum Theory, one finds that

$$\vec{\mu} = \frac{e}{2mc} g \vec{S}, \tag{18.30}$$

where m is the electron mass and $g = 2$. Actually, for the electron this quantity has been measured with great accuracy, with the result that $g \approx 2.0023193043617$. The small discrepancy is due to the fact that a real electron is the source of a field, not considered in Dirac's theory, and as such, it interacts with virtual electron–positron pairs. Quantum Electrodynamics is a field theory, taking into account the interaction, and its result agrees extremely well with the data. For the proton, which is about 1800 times more massive and has internal structure and strong interactions, $g \approx 5$, and μ is much smaller. The neutron has μ comparable to that of the proton.

In a magnetic field B, the dipole energy is given by (18.3) like in Classical Physics. The Schrödinger–Pauli equation reads as

$$i\hbar \frac{\partial \psi}{\partial t} = H\psi. \tag{18.31}$$

For a particle with charge e and spin 1/2 in a magnetic field,

$$H = \begin{pmatrix} \frac{(\vec{p} - \frac{e}{c}\vec{A})^2}{2m} & 0 \\ 0 & \frac{(\vec{p} - \frac{e}{c}\vec{A})^2}{2m} \end{pmatrix} - \frac{e}{2mc} \mathrm{g} \vec{S} \cdot \vec{B}, \tag{18.32}$$

that is,

$$\begin{pmatrix} \frac{(\vec{p} - \frac{e}{c}\vec{A})^2}{2m} - \frac{e\hbar}{2mc}\frac{g}{2}B & 0 \\ 0 & \frac{(\vec{p} - \frac{e}{c}\vec{A})^2}{2m} + \frac{e\hbar}{2mc}\frac{g}{2}B \end{pmatrix}. \tag{18.33}$$

The magnetic term $\frac{e}{2mc} g \vec{S} \cdot \vec{B}$ alone has spin up eigenstates $|\alpha\rangle$ with eigenvalue $\mu_B B$ and spin down eigenstates $|\beta\rangle$ with $-\mu_B B$, where μ_B is the *Bohr–Procopiu magneton*; for $e < 0$ the solutions are exchanged; so, for an electron in a magnetic field parallel to the z axis, $|\beta\rangle$ is the ground state. Approximating g with 2,

$$\mu_B = \frac{e\hbar}{2mc}. \tag{18.34}$$

The two levels are separated by the *Larmor frequency*,

$$\omega_L = \frac{eB}{mc}.$$

Including the kinetic term only, one finds the Landau levels (Sect. 17.1); the spin term replaces each level by two sublevels; since $g \sim 2$, the spin down adds $\frac{eB}{2mc}$ to the energy and the spin up subtracts the same. Thus, the two magnetic levels are separated by $\sim \hbar\omega_L$.

The direct transition between the spin sublevels can be done by absorbing or emitting a photon, in a magnetic dipole transition; this is EPR (electron paramagnetic resonance, also called electron spin resonance), or NMR (nuclear spin resonance). For a proton the ground state is α, and the frequency is three orders of magnitude lower.

Problem 18.2 Rotate the spinor $|\alpha\rangle$ around the x (a) by an angle $\varphi = \pi/2$ and (b) by an angle $\varphi = \pi$. Interpret the results physically.

Solution 18.2 The rotation operator by φ around x is

$$R_\varphi = \exp\left[\frac{-i\varphi}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right].$$

Separating odd and even powers in the expansion of $e^{ix} = \cos(x) + i\sin(x)$, one finds

$$R_\varphi = \sum_{\text{evenn}} \frac{1}{n!}\left(\frac{-i\varphi}{2}\right)^n + \sigma_x \sum_{\text{oddn}} \frac{1}{n!}\left(\frac{-i\varphi}{2}\right)^n.$$

$$\sum_{\text{npari}} \frac{(ix)^n}{n!} = Ree^{ix} = \cos(x) \qquad \sum_{\text{ndispari}} \frac{(ix)^n}{n!} = iIme^{ix} = i\sin(x).$$

Therefore,

$$R_\varphi = \begin{pmatrix} \cos\left(\frac{\varphi}{2}\right) & -i\sin\left(\frac{\varphi}{2}\right) \\ -i\sin\left(\frac{\varphi}{2}\right) & \cos\left(\frac{\varphi}{2}\right) \end{pmatrix}.$$

Note that $R_\varphi^\dagger = R_\varphi^{-1}$. One finds (a)

$$R_{\pi/2}|\alpha\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

This is an eigenstate (spin down) of S_y. (b)
$R_\pi|\alpha\rangle = -i|\beta\rangle$. This is in line with one's intuition.

Problem 18.3 Verify the identity

$$\overrightarrow{j}(1)\cdot\overrightarrow{j}(2)=j_z(1)j_z(2)+\frac{j_+(1)j_-(2)+j_-(1)j_+(2)}{2},$$

where $\overrightarrow{j}(1)$, $\overrightarrow{j}(2)$ are angular momenta of different particles with $[\overrightarrow{j}(1), \overrightarrow{j}(2)]_-=0$.

Solution 18.3 Just substitute the definitions.

18.4 Quantum "Sum" of Angular Momenta

The isotropy of space implies that the Hamiltonian must commute with all rotations, and with their generators. As a result, the total angular momentum of any isolated system is interesting, since it is conserved. Consider a system made up of two parts with angular momenta $\boldsymbol{j}_1$ and $\boldsymbol{j}_2$ that depend on different variables and therefore commute:

$$[\boldsymbol{j}_1,\boldsymbol{j}_2]_-=0; \tag{18.35}$$

the angular momentum operator of the system is the sum

$$\boldsymbol{j}=\boldsymbol{j}_1+\boldsymbol{j}_2, \tag{18.36}$$

as in Classical Physics. In interesting cases, $\boldsymbol{j}_1$ and $\boldsymbol{j}_2$ are mixed by some mutual interaction, while $\boldsymbol{j}$ commutes with the Hamiltonian and is conserved. The mathematical properties of angular momentum tell us how the $\boldsymbol{j}$ eigenstates can be expanded.

Now, $\boldsymbol{j}$ is really an angular momentum, since Eqs. (18.35) and (18.36) imply

$$\boldsymbol{j}\wedge\boldsymbol{j}=i\hbar\boldsymbol{j}. \tag{18.37}$$

To start with, we have a basis of eigenstates for each part:

$$\begin{aligned} j_1^2|j_1m_1\rangle &= j_1(j_1+1)\hbar^2|j_1m_1\rangle,\ j_{1z}|j_1m_1\rangle=m_1\hbar|j_1m_1\rangle\\ j_2^2|j_2m_2\rangle &= j_2(j_2+1)\hbar^2|j_2m_2\rangle,\ j_{2z}|j_2m_2\rangle=m_2\hbar|j_2m_2\rangle. \end{aligned}$$

The direct product basis is a basis for the compound system:

$$|j_1m_1j_2m_2\rangle\equiv|j_1m_1\rangle|j_2m_2\rangle. \tag{18.38}$$

For all the $(2j_1+1)(2j_2+1)$ basis vectors, we already know the eigenvalue m_j of j_z:

$$j_z|j_1m_1j_2m_2\rangle=(m_1+m_2)\hbar|j_1m_1j_2m_2\rangle=m_j\hbar|j_1m_1j_2m_2\rangle.$$

A complication arises, since one cannot (generally) assign j, m_1 and m_2 simultaneously. In fact,

$$\boldsymbol{j}^2 = \boldsymbol{j}_1^2 + \boldsymbol{j}_2^2 + 2\,\boldsymbol{j}_1 \cdot \boldsymbol{j}_2; \tag{18.39}$$

here, the squares do commute with the components, but the last term $\boldsymbol{j}_1 \cdot \boldsymbol{j}_2$ fails to commute:

$$\left[\boldsymbol{j}^2, j_{1z}\right]_- = 2\left[\boldsymbol{j}_1 \cdot \boldsymbol{j}_2, j_{1z}\right]_- = 2\left[j_{1x} j_{2x} + j_{1y} j_{2y}, j_{1z}\right]_- \tag{18.40}$$

does not vanish. We can label states by j, j_1, j_2, and of course $m = m_1 + m_2$ since we recall from Chap. 13 that the square of an angular momentum and one of its components exist simultaneously.

Problem 18.4 Compute the commutator (18.40).

Solution 18.4 One finds $2i\hbar(\boldsymbol{j}_1 \wedge \boldsymbol{j}_2)_z$.

Sum of Two Spins 1/2

In the ground state of the H atom,[6] the orbital angular momentum is $l = 0$; therefore, there is no spin–orbit interaction. But there are two 1/2 spins to sum, one electron and a proton spin. The nuclear angular momentum would be conserved separately, if we could neglect the weak magnetic dipole coupling with the electron angular momentum. We shall see below that it is very good for Astrophysics that because of the magnetic dipole–dipole interaction, neither angular momentum is conserved (while of course the resultant gives good quantum numbers j, m_j). Let us denote by $\alpha(1), \beta(1)$ the eigenstates of the electron S_z and by $\alpha(2), \beta(2)$ those of the nucleus; the direct product of the two bases gives the four states in Table 18.1.

We can label all four states with the total angular momentum component m_j by simply summing the individual components, since the operators commute; we can also put a total $j = 1$ label on the first and last lines. Indeed, $j = 0$ is excluded because it must give $m_j = 0$, while higher values are excluded since they should also give $m_j > 1$ components that do not exist. The four-dimensional Hilbert space has a three-dimensional subspace $j = 1$, called the *triplet*, and the component $|j = 1, m_j = 0\rangle$, called the *singlet*, must be a combination of the two states with $m_j = 0$. To separate

Table 18.1 Spin states of H

Basis states	m_j	j
$\|\alpha(1)\alpha(2)\rangle$	1	1
$\|\alpha(1)\beta(2)\rangle,\ \|\beta(1)\alpha(2)\rangle$	0	Mixed 1, 0
$\|\beta(1)\beta(2)\rangle$	−1	1

[6] Specifically, I refer to Protium ^1H; the Deuterium ^2H nucleus has a neutron and a proton and spin 1, and the unstable Tritium ^2H has spin 1/2, too.

the two components, one can start from $|\alpha(1)\alpha(2)\rangle$ and lower the z component by acting with $j^- = S^-(1) + S^-(2)$.

In the example of Table 18.1, using the shift operators (18.10) $S^+ = \hbar \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $S^- = \hbar \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, the required shift operator is $\boldsymbol{j}^- = S_-(1) + S_-(2)$.

Since $S^-(1)|\alpha(1)\alpha(2)\rangle = |\beta(1)\alpha(2)\rangle$, $S^-(2)|\alpha(1)\alpha(2)\rangle = |\alpha(1)\beta(2)\rangle$, we obtain at once the $m_j = 0$ component of the triplet

$$|j = 1, m_j = 0\rangle = \frac{\alpha(1)\beta(2)\rangle + \beta(1)\alpha(2)\rangle}{\sqrt{2}}. \tag{18.41}$$

The orthogonal combination must be the singlet

$$|j = 0, m_j = 0\rangle = \frac{\alpha(1)\beta(2)\rangle - \beta(1)\alpha(2)\rangle}{\sqrt{2}}. \tag{18.42}$$

To verify that these are indeed eigenstates of $j^2 = S_1^2 + S_2^2 + 2\boldsymbol{S}_1\boldsymbol{S}_2$, one can start with $S_1^2 = S_2^2 = \frac{3}{4}\hbar^2$ and note that $\boldsymbol{S}_1\boldsymbol{S}_2 = S_{1z}S_{2z} + S_{1x}S_{2x} + S_{1y}S_{2y}$; hence $j^2 = \frac{3}{2}\hbar^2 + 2[S_{1z}S_{2z} + \frac{1}{2}(S_1^+S_2^- + S_1^-S_2^+)]$. This makes the verification immediate. The separation between the singlet ground state and the triplet excited state is small (5.9×10^{-6} eV) and belongs to the so-called hyperfine structure of H. It corresponds to a 21 cm wavelength line (at 1420 MHz) and it is very important in Astrophysics, since it is weakly absorbed by dust clouds, and so allows us to see the atomic H clouds throughout the Milky Way and beyond.

18.4.1 Clebsch–Gordan Coefficients

Suppose we have a system consisting of two parts with angular momenta j_1 and j_2, which would be conserved in the absence of a mutual interaction. From the separate bases $|j_1, m_1\rangle$ and $|j_2, m_2\rangle$, one obtains the so-called *direct product basis* for the whole system, with elements:

$$\{|j_1 m_1 j_2 m_2\rangle\}, \qquad m_1 \in (-j_1, j_1), m_2 \in (-j_2, j_2). \tag{18.43}$$

But since the total angular momentum is conserved, we may wish to go to a new basis $|J, M_j\rangle$ where $\hat{J} = \hat{j}_1 + \hat{j}_2$ and M is the eigenvalue of $J_z = j_{1z} + j_{2z}$. The maximum value of M_j is $M_j^{max} = j_1 + j_2$ and the minimum is $M_j^{min} = -M_j^{max}$. Therefore, the space spanned by the direct product basis includes the space spanned by the basis of $J = j_1 + j_2$, and this is the maximum J that can occur (since larger J would imply larger M_j). Consequently, by taking linear combinations of the product basis vectors, one can form a basis for the subspace $J = j_1 + j_2$; the number of linear combinations

for this J is $2(j_1 + j_2) + 1$. As we know, these states are labeled by J and M_j, but not all of them have sharp m_1 and m_2. In fact only the maximum M_j and the minimum M_j have such sharp values. Instead, $M_j = j_1 + j_2 - 1$ corresponds to two different members of the direct product basis, namely, $|j_1, m_1 = j_1, j_2, m_2 = j_2 - 1\rangle$ and $|j_1 m_1 = j_1 - 1, j_2 m_2 = j_2\rangle$. From these vectors, we know how to make one linear combination belonging to $J = j_1 + j_2$: we must just apply the shift operator $\hat{J}^- = \hat{j}_1^- + \hat{j}_2^-$ to $|J, M = J\rangle$, getting $|J, M = J - 1\rangle$. The orthogonal combination must then belong to $J = j_1 + j_2 - 1$ and $M_j = j_1 + j_2 - 1$. Now, $J = j_1 + j_2 - 1$ must occur only once. In this way, one finds all the values from $j = j_1 + j_2$ to $j = |j_1 - j_2|$ and all appear once. Indeed, one can prove the following identity:

$$\sum_{j=|j_1-j_2|}^{j_1+j_2} (2j + 1) = (2j_1 + 1)(2j_2 + 1). \tag{18.44}$$

Thus, in general, one can go from the direct product basis to the total angular momentum basis $|J M_j\rangle$ by a unitary transformation

$$|j m_j\rangle = \sum_{m_1=-j_1}^{j_1} \sum_{m_2=-j_2}^{j_2} |j_1 m_1 j_2 m_2\rangle \langle j_1 m_1 j_2 m_2 | j m_j\rangle. \tag{18.45}$$

One can do the reverse transformation by the inverse unitary transformation. The *Clebsch–Gordan coefficients* $\langle j_1 m_1 j_2 m_2 | j m_j\rangle$ can be found by using the shift operator and orthogonalization as we have just seen. This is just the procedure that we used to derive equations (18.41), (18.42). There are extensive tables of Clebsch–Gordan coefficients, that are also readily obtained by computer systems like Mathematica. Here is another example.

Sum of $L = 1$ and a Spin 1/2

In atomic Physics, the electron orbital momentum is coupled to the electron spin momentum by the relativistic spin–orbit interaction arising in Dirac's theory. Indeed, the relativistic many-body theory is mandatory in atomic calculations. However, when relativistic effects are small, the second-order corrections in v/c are a fair approximation. The spin–orbit interaction

$$H' = -\frac{e\hbar}{4m^2c^2} \boldsymbol{\sigma}.(\boldsymbol{E} \wedge \boldsymbol{p}) \tag{18.46}$$

is prominent among the corrections arising from Dirac's theory since it can lift degeneracies.

To find the angular momentum of an electron in a $2p$ orbital, we must combine $l = 1$ with the spin ($3 \times 2 = 6$ states). One can denote the states by $|m_l, m_s\rangle$.

The maximum and the minimum m_j must belong to $j = \frac{3}{2}$ (it cannot be less, because $m_j = \frac{3}{2}$ exists, but it cannot be more, because otherwise higher m_j should appear). But $j = \frac{3}{2}$ means four states, and the other two must belong to $j = \frac{1}{2}$.

Table 18.2 Angular momentum of a 2p electron

States	m_j	j
$\|1, \frac{1}{2}\rangle$	$\frac{3}{2}$	$\frac{3}{2}$
$\|1, -\frac{1}{2}\rangle,\ \|0, \frac{1}{2}\rangle$	$\frac{1}{2}$	Mixed $\frac{3}{2}, \frac{1}{2}$
$\|0, -\frac{1}{2}\rangle,\ \|-1, \frac{1}{2}\rangle$	$-\frac{1}{2}$	Mixed $\frac{3}{2}, \frac{1}{2}$
$\|-1 - \frac{1}{2}\rangle$	$-\frac{3}{2}$	$\frac{3}{2}$

We know that (Table 18.2)

$$|\frac{3}{2}, \frac{3}{2}\rangle = |1, 1\rangle|\frac{1}{2}\rangle.$$

The right combination for $j = \frac{3}{2}, m_j = \frac{1}{2}$ can be found by acting on the state $m_j = \frac{3}{2}$ with

$$j_- = L_- + s_-.$$

Using the general formula $L_\pm|l, m\rangle = \sqrt{l(l+1) - m(m \pm 1)}\hbar|l, m \pm 1\rangle$, and normalizing,

$$|\frac{3}{2}, \frac{1}{2}\rangle = \sqrt{\frac{2}{3}}|1, 0\rangle|\frac{1}{2}\rangle + \frac{1}{\sqrt{3}}|1, 1\rangle|\frac{-1}{2}\rangle. \tag{18.47}$$

The orthogonal combination is

$$|\frac{1}{2}, \frac{1}{2}\rangle = \frac{1}{\sqrt{3}}|1, 0\rangle|\frac{1}{2}\rangle - \sqrt{\frac{2}{3}}|1, 1\rangle|\frac{-1}{2}\rangle. \tag{18.48}$$

From (18.47) we deduce that

$$\langle 1, 0, \frac{1}{2}, \frac{1}{2}|\frac{3}{2}, \frac{1}{2}\rangle = \sqrt{\frac{2}{3}},\ \langle 1, 1, \frac{1}{2}, -\frac{1}{2}|\frac{3}{2}, \frac{1}{2}\rangle = \sqrt{\frac{1}{3}},$$

and from (18.48) it follows that

$$\langle 1, 0, \frac{1}{2}, \frac{1}{2}|\frac{1}{2}, \frac{1}{2}\rangle = \sqrt{\frac{1}{3}},\ \langle 1, 1, \frac{1}{2}, -\frac{1}{2}|\frac{1}{2}, \frac{1}{2}\rangle = -\sqrt{\frac{2}{3}}.$$

Thus, one can do very well without a table of Clebsch–Gordan coefficients.

Table 18.3 The Clebsch–Gordan coefficients for $j_1 = j_2 = \frac{1}{2}$

m_1	m_2	j	$\langle j_1 j_2 m_1 m_2 \vert jm = m_1 + m_2\rangle$
$\frac{1}{2}$	$\frac{1}{2}$	1	1
$\frac{1}{2}$	$-\frac{1}{2}$	1	$\frac{1}{\sqrt{2}}$
$\frac{1}{2}$	$-\frac{1}{2}$	0	$\frac{1}{\sqrt{2}}$
$-\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{\sqrt{2}}$
$-\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{\sqrt{2}}$
$-\frac{1}{2}$	$-\frac{1}{2}$	1	1

Table 18.4 The i Clebsch–Gordan coefficients for $j_1 = 1$, $j_2 = \frac{1}{2}$

m_1	m_2	j	$\langle j_1 j_2 m_1 m_2 \vert jm = m_1 + m_2\rangle$
1	$\frac{1}{2}$	$\frac{3}{2}$	1
1	$-\frac{1}{2}$	$\frac{3}{2}$	$\sqrt{\frac{1}{3}}$
1	$-\frac{1}{2}$	$\frac{1}{2}$	$\sqrt{\frac{2}{3}}$
0	$\frac{1}{2}$	$\frac{3}{2}$	$\sqrt{\frac{2}{3}}$
0	$\frac{1}{2}$	$\frac{1}{2}$	$-\sqrt{\frac{1}{3}}$
0	$-\frac{1}{2}$	$\frac{3}{2}$	$\sqrt{\frac{2}{3}}$
0	$-\frac{1}{2}$	$\frac{1}{2}$	$\sqrt{\frac{1}{3}}$
-1	$\frac{1}{2}$	$\frac{3}{2}$	$\sqrt{\frac{1}{3}}$
-1	$\frac{1}{2}$	$\frac{1}{2}$	$-\sqrt{\frac{2}{3}}$
-1	$-\frac{1}{2}$	$\frac{3}{2}$	1

Action of j^2 on the LS Basis

In order to evaluate the matrix elements and to verify if a wave function is eigenstate of j^2, one can combine the identity $j^2 = L^2 + S^2 + 2\boldsymbol{L}\cdot\boldsymbol{S}$ with

$$\overrightarrow{L}\cdot\overrightarrow{S} = L_z S_z + \frac{1}{2}\left(L_+ S_- + L_- S_+\right),$$

recalling that (Tables 18.3 and 18.4)

$$\langle l_1 m_1 | L_\pm | l_2 m_2\rangle = \hbar\sqrt{l_1(l_1+1) - m_2(m_2 \pm 1)}\,\delta_{l_1,l_2}\delta_{m_1\pm 1,m_2}.$$

When summing more than two angular momenta, one can choose an arbitrary order and combine the first two, then the resulting sums with the third, and so on.

The final pattern of possible total angular momenta does not depend on the arbitrary order; in fact, it is an objective physical reality. However, when several multiplets with the same j appear, their bases can get mixed by changing the order.

18.5 Photon Spin and Quantum Cryptography

There is an obvious difference between electron spin and photon spin: the photon, unlike the electron, is a vector particle, since its wave function is the four-potential A_μ. In the case of a photon going up along z, we can take an orthogonal basis $(|x\rangle, |y\rangle)$ of linearly polarized states. The spin states $\sigma_z = \pm 1$ correspond to the combinations $\frac{|x\rangle \pm i|y\rangle}{\sqrt{2}}$. A field $\overrightarrow{E}$ polarized at 45° is like a vector that points at 45° $\frac{|x\rangle + |y\rangle}{\sqrt{2}}$ and is an equal mixture of both polarizations. This quantum property has an important application in cryptography.

Everybody keeps secrets. There is information that must remain secret for a long time, like passwords, bank coordinates or the controls of a commercial satellite—not to mention classified military and security files. For thousands of years, there has been a competition between those who wanted to encrypt secrets with increasingly ingenious codes and those who wanted to decipher them. The secret code that Julius Caesar used was the Caesar cipher, a cryptography system based on the alphabet, but with a numerical key. For example, if the key was 3, each letter of the original message was replaced with the letter that was three positions ahead in the alphabet. So, the word ROMA became URPD. The Allies made great progress after breaking the code of the Enigma machines of the Wehrmacht at the beginning of World War II.

With the advent of Quantum Mechanics, the competition is now being won by those who want to keep secrets. Various devices to achieve cryptography using the uncertainty principle have come to the industrial level and are sold by various commercial companies. Here is the principle of operation.

The basic problem of cryptography is well known: two people, let us say Alice and Bob, need to exchange secret information, using a public channel such as a telephone line or radio; people who have interest in the breaking the code can receive signals, but cannot possibly understand them without a cryptographic key, which may be represented by a secret number. If the number has many digits, the secret can be effectively kept. But first, Bob and Alice must share this number. The essential problem that Alice and Bob must solve is this: how to exchange the key while avoiding it also being received by the eavesdropper who has access to the public channel.

Charles Bennet of IBM ed Gilles Brassard of Montreal University developed a method based on Quantum Mechanics to do this in a safe way. The signal is transmitted as a succession of bits that can be 1 or 0. First, Alice and Bob agree that a photon polarized "horizontally" is 1 and one polarized "vertically" means 0. In order to transmit and receive photons, both have parallel polarizing filters that can

emit the light and read the polarizations. The filters act like projectors $|x\rangle\langle x|$, which beep if the signal is polarized along x. However, for the present purpose, Alice and Bob also need a second filter rotated 45° compared to the first. So, indicated with a dash of the polarization plane, the bit 1 can travel as − or \, while bit 0 can travel as | or /.

If the filter used to receive is the same as that used to transmit, the bit is read without errors, but if, for example, a signal emitted as | is picked up with the inclined filter, half of the time, it will be interpreted as −.

The transmission of the key is done as follows. Alice sends a random bit sequence, even with randomly choosing the filter for sending them, and keeping track each time about what she does. For each incoming bit, Bob randomly chooses which filter to use to read it and makes a note of the filter used for the reading. After the end of the photon transmission, Bob communicates to Alice, through the public channel, the sequence of filters used in the readings (obviously without communicating the sequence of readings). Next, Alice can reveal, using the same channel, which filters were used correctly. Then, deleting the photons read by the wrong filter, both Alice and Bob have obtained a sequence of bits that constitute the secret key.

A possible eavesdropper can receive the photons, but Quantum Mechanics prevents him from doing what classically would be possible, i.e., he cannot observe the state of polarization without the risk of changing it. Choosing between − and | means choosing between two orthogonal states, like the states α and β of a spin. When the eavesdropper receives the photon with the wrong filter, he not only reads the polarization incorrectly, but can also change the polarization. This is similar to the effect of the second Stern–Gerlach filter in Fig. 18.2. The errors introduced in this way will eventually reveal his presence. Indeed, randomly choosing a small number of bits among those making up the message, Alice and Bob can verify on the public channel whether they arrived unchanged or not; in the latter case, the presence of the eavesdropper will be unmasked.

18.5.1 Rabi Model

As an example of the nontrivial behavior of quantum time-dependent systems, consider the Rabi model, which is fundamental in Quantum Optics.[7] It represents a spin in a magnetic field or a two-level atom, with separation ω_0 between the ground state level E_g and the excited state E_e.

Suppose that the model is initially in the ground state $|g\rangle$; then, it starts interacting with a radiation field whose frequency ω is comparable to ω_0. The difference $\Delta = \omega - \omega_0$ is called detuning. The Rabi Hamiltonian is

$$H = -\frac{\hbar\omega_0}{2}\sigma_z - \tau\cos(\omega t)\sigma_x. \tag{18.49}$$

[7] Isidor Isaac Rabi (1898–1988) was a U.S. Physicist, who received the Nobel Prize in 1944.

Perhaps, one might expect that the probability P of finding the atom in the excited state $|e\rangle$ should vanish through energy conservation whenever $\Delta \neq 0$. However this is not true, because H depends on the time, and both levels become dressed with light. Let $E_e = \frac{\hbar\omega_0}{2} = -E_g$, and let us write the wave function as

$$\psi(t) = a_g|g\rangle \exp\left(\frac{-iE_g t}{\hbar}\right) + a_e|e\rangle \exp\left(\frac{-iE_e t}{\hbar}\right); \tag{18.50}$$

the Schrödinger equation yields

$$\begin{aligned} \dot{a}_g &= \tfrac{i}{\hbar}\tau \cos(\omega t) \exp\left(\tfrac{-i\omega_0 t}{\hbar}\right) a_e \\ \dot{a}_e &= \tfrac{i}{\hbar}\tau \cos(\omega t) \exp\left(\tfrac{i\omega_0 t}{\hbar}\right) a_g. \end{aligned} \tag{18.51}$$

This is readily solved exactly, but to allow for a simple analytic formula, we can neglect the quickly rotating terms in $\omega + \omega_0$, while keeping the slow ones in $\omega - \omega_0$. This is known as the Rotating Wave Approximation. Then, we obtain the easily solvable equations

$$\begin{aligned} \dot{a}_g &= \tfrac{i}{2\hbar}\tau \exp(\tfrac{i(\omega-\omega_0)t}{\hbar}) a_e \\ \dot{a}_e &= \tfrac{i}{2\hbar}\tau \exp(\tfrac{-i(\omega-\omega_0)t}{\hbar}) a_g. \end{aligned} \tag{18.52}$$

Now it is elementary to find that

$$\alpha_e(t) = \frac{i\tau}{\hbar\Omega} \exp\left(\frac{it}{2}\Delta\right) \sin\left(\frac{\Omega t}{2}\right); \tag{18.53}$$

the Rabi frequency Ω defined by

$$\hbar\Omega = \sqrt{\tau^2 + (\hbar\Delta)^2}. \tag{18.54}$$

For $\Delta = 0$, the atom jumps up and down with the frequency $\frac{\tau}{\hbar}$. Increasing the detuning makes the Rabi oscillations more frequent, while the system spends less and less time in the excited state.

18.6 Space–Time Symmetries of Pauli Electrons

Stereoisomers that are mirror images of each other are called enantiomers. For example, Lactic Acid $(CH_3)(CHOH) - COOH$ has enantiomer forms. It is clear that the molecular spin-orbitals in a one-electron (Hartree–Fock) description must be simply related. Let H'_{SO} denote the spin–orbit interaction and (in obvious notation)

$$H\psi_{k,\lambda} = (\frac{p^2}{2m} + V(x) + H'_{SO})\psi_{k,\lambda} = \epsilon_{k,\lambda}\psi_{k,\lambda}, \tag{18.55}$$

with λ a spin index, be the Pauli equation for an electron state in one of the enantiomers; then how should we describe the correspondent state in the mirror image? It is the same problem, but with V replaced by $V^{(i)}(x) = PV(x) = V(-x)$, where P is the Parity operator, while the spin–orbit operator changes sign twice and remains the same. So the one-electron eigenvalue $\epsilon_{k,\lambda}$ is obviously the same for both enantiomers.[8]

Next, we need to show that time reversal is also an important symmetry. Suppose we can solve the spinless Schrödinger equation with no magnetic field (real Hamiltonian)

$$i\hbar\frac{\partial\phi(t)}{\partial t} = H(t)\phi(t), \tag{18.56}$$

with real $H(t)$. How can we use the solution to solve the time-reversed dynamics

$$i\hbar\frac{\partial\phi'}{\partial t} = H(-t)\phi'(t)? \tag{18.57}$$

Set $t' = -t$. We can find an operator T such that $\phi'(t') = T\phi(t)$ and time-reversed operators are given by $\hat{A}' = T\hat{A}T^{-1}$. To this end, introduce the Kramers operator K that takes the complex conjugate: $K\phi = \phi^*$. Then,

$Ki\hbar\frac{\partial}{\partial t} = i\hbar\frac{\partial}{\partial t'}$ and for any two spinors ϕ and ψ , $\langle K\phi|K\psi\rangle = \langle\psi|\phi\rangle$. Thus, the time-reversed $\overrightarrow{p}$ operator is just $-\overrightarrow{p}$. Applying K to the Schrödinger equation $i\hbar\frac{\partial\phi(\tau)}{\partial\tau} = H(\tau)\phi(\tau)$, and using the fact that H is real, one finds $-i\hbar\frac{\partial\phi^*(\tau)}{\partial\tau} = H(\tau)\phi^*(\tau)$; now, setting $\tau = -t$, we obtain $i\hbar\frac{\partial\phi^*(-t)}{\partial t} = H(-t)\phi^*(-t)$. Comparing with the time-reversed dynamics $i\hbar\frac{\partial\phi'(t)}{\partial t} = H(-t)\phi'(t)$, we conclude that

$$\phi'(t) = \phi^*(-t). \tag{18.58}$$

Introducing spin and magnetic field makes things more complicated. Starting with the Pauli equation with a magnetic field,

$$i\hbar\frac{\partial\psi}{\partial t} = [H_0 + \lambda\overrightarrow{\sigma}.\overrightarrow{B}(t)]\psi, \tag{18.59}$$

where $\lambda = \frac{e\hbar}{2mc}$, H_0 is the spin-independent (that is, spin diagonal) part of the Hamiltonian,

$$H_0(t) = \frac{1}{2m}[p^2 + (\frac{eA(t)}{c})^2 - \frac{2e\overrightarrow{A}(t).\overrightarrow{p}}{c}] + V(t), \tag{18.60}$$

[8] So, in crystals with space inversion symmetry all electron states are degenerate.

the time-reversed dynamics is given by

$$i\hbar\frac{\partial\psi'}{\partial t} = [H_0'(-t) + \lambda\overrightarrow{\sigma}.\overrightarrow{B}'(-t)]\psi'; \tag{18.61}$$

here, the primes are needed: indeed, the currents change sign under time reversal, hence the vector potential and the magnetic field also change sign. If we wish to write down the reverse dynamics in terms of the original fields, $\overrightarrow{A}' = -\overrightarrow{A}$, $\overrightarrow{B}' = -\overrightarrow{B}$. Thus,

$$\psi'(t) = -i\sigma_2\psi^*(-t). \tag{18.62}$$

So, to obtain the time-reversed dynamics, not only $t \to -t$, not only K takes the complex conjugate, but the action of σ_2 also reverses the spin. In other terms, the time reversal operator for Pauli spins is

$$T = -i\sigma_2 K. \tag{18.63}$$

Note that the Pauli spinors have the property:

$$\langle K\phi|K\psi\rangle = \langle\psi|\phi\rangle. \tag{18.64}$$

Moreover $T^2 = -1$, so $T^{-1} = -T$. Instead, the spin–orbit interaction is time reversal invariant. When there is no magnetic field (and no time dependence), T is a symmetry of the Pauli Hamiltonian and implies degeneracy of $K\psi$ and ψ. This is called the Kramers degeneracy. The above properties are important in the theory of crystals. Some have both symmetries, parity P and time reversal T; hence they have the joint symmetry $C = PT$ which is called Conjugation. When C is a symmetry, there is spin degeneracy at any wave vector even in the presence of the spin–orbit interaction.

18.7 Path Integral Formulation of Quantum Mechanics. Part 2: Spin

In Chap. 15, the Feynman path integral is presented for a spinless particle. Here we are in position to present in an intuitive way[9] the path integral approach for a spin S interacting with a magnetic field B with a coupling of the form:

$$H = \hat{B}\cdot\hat{S}. \tag{18.65}$$

[9] For a more in depth treatment of the related mathematical aspects the reader may consult the book "Condensed Matter Field Theory" by Alexander Altland and Ben Simons, Cambridge.

To this end, we use the Euler angle representation

$$g(\phi, \theta, \psi) = e^{-i\phi S_3} e^{-i\phi S_2} e^{-i\psi S_3} | \uparrow\rangle$$

defined above. Basically, the idea of the path integral is the same as in Sect. 15.7, namely, the time interval (0,t) is sliced into very thin slices, complete sets are inserted between them, the thin slices allow some simplification and the full evolution is finally obtained in the form of infinite nested integrals. For a realistic treatment of a real spin 1/2 particle, the complete sets should involve all the space and spin degrees of freedom, with the space part which interacts via the vector potential while the spin interacts with the field.[10] However here I wish to focus on the main points with a minimum of formalism. We shall see that a treatment just involving the spin only is interesting by its own right.

As in Sect. 15.7, the amplitude $A(t) = \langle a|U(0,t)|b\rangle$ to go from state a at time t = 0 to a state b at time t is obtained by slicing the time interval from t_a to t_b in intervals of very short duration ϵ. This breaks the matrix element into a product of factors, each representing the evolution over a short time ϵ. Thus, letting dg_i denote the change of g in the ith interval, we must integrate in $\prod_{i=0}^{N} dg_i \equiv Dg$, with a very large N. Next, we have to insert a suitably chosen resolution of identity between the factors. This is a crucial step in the construction of the path integral.

In each infinitesimal interval lasting ϵ one can write $e^{-\epsilon \vec{B}\,.\,\vec{S}} \sim 1 - \epsilon \vec{B}\,.\,\vec{S}$. Thus,

$$\langle g_{i+1}|e^{-\epsilon \vec{B}\,.\,\vec{S}}|g_i|\rangle \sim \langle g_{i+1}|g_i\rangle - \epsilon\langle g_{i+1}|\vec{B}\,.\,\vec{S}\,|g_i\rangle.$$

To proceed we need a trick. First, we add $1 - \langle g_i|g_i\rangle$, which is nothing; next, we interpret the result as the beginning of the expansion of

$$\exp(\langle g_i + 1|g_i\rangle - \langle g_i|g_i\rangle - \epsilon\langle g_{i+1}|\vec{B}\,\vec{S}\,|g_i\rangle).$$

Thus, we obtain the result:

$$A(t) = \int Dg \exp[-d\tau(-\langle\partial_\tau g|g\rangle + \langle g|\vec{B}\,.\,\vec{S}\,|g\rangle)] \tag{18.66}$$

which can be interpreted as a sum over all possible paths $g(\tau)$ and is the path integral for spin.

Reference

Beth, R.A.: Phys. Rev. **5**, 115 (1936)

[10] We know from Sect. 15.5 that the interaction of the spin dipole with the magnetic field cannot be represented by a vector potential expression valid in all space, for topological reasons.

Chapter 19
Systems of Particles and Second Quantization

The study of many-body systems requires new concepts in dramatic contrast with Classical Physics. Quantum particles can be in an entangled state, and like particles must be entangled because of a permutation symmetry of the wave function. All that has enormous consequences in all of Physics.

19.1 The Pauli Principle

We need to extend the quantum mechanical theory to the case of N particles; separately, the particles would be described by a Schrödinger equation and the Maxwell equations if they are photons or the Pauli equation, for electrons. As in the classical case, the Hamiltonian of the system will be the sum of those of the particles plus (possibly) an interaction term.

The wave function $\Psi(1, 2, \ldots, N)$ depends on all the orbital and spin degrees of freedom of the particles, so it is a spinor in the spin space of each particle. Consequently, the scalar product of two wave functions Ψ and Φ can be written symbolically as

$$\langle \Phi | \Psi \rangle = \int \Phi^* \Psi d1 \ldots dN, \tag{19.1}$$

where, besides the integration over the coordinates, one must understand the scalar products of the corresponding spins, for particles that possess one. In this way, one can define normalized states.

For identical particles, a new important ingredient comes into play, which is called the permutation symmetry. Consider, for instance, two electrons, and let A denote the spinor of electron 1 and B denote the spinor of electron 2. Let the operator $P(1, 2)$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_19

exchange two electrons (coordinates and spins), putting 1 in B and 2 in A. It is clear that the physical state does not change. There can exist no way to distinguish one electron from another. They are unlike billiard balls, which can be marked with a chalk. In the wave function, they are marked 1 and 2, so $P(1,2)$ does a mathematical operation. Since $P(1,2)^2 = 1$, $P(1,2)$ has eigenvalues of ± 1, one could argue that sometimes we get 1 and sometimes -1. That's reasonable. But instead, we get always -1, when exchanging two electrons, or two protons, or two identical particles having half-integer spins (even 3/2, 5/2 and so on). These are called Fermions. On the other hand, the Bosons (integer spin particles) always have $P(1,2) = 1$. In summary, for two particles:

$$P(1,2)\Psi(1,2) = \Psi(2,1) = \begin{cases} -\Psi(1,2) & \text{Fermions,} \\ \Psi(1,2) & \text{Bosons.} \end{cases} \tag{19.2}$$

At the fundamental level, Bosons are related to fields and Fermions to matter. Pions (spin 0), Photons and the vector particles $W^{\pm}$, Z^0 (spin 1) are Bosons, like Gravitons (spin 2, provided they exist, who knows). Instead, Electrons, Neutrinos, Quarks and nucleons are 1/2 spin Fermions. On the other hand, composite systems like nuclei and atoms may have integer or half-integer spins in their ground states and will behave accordingly at low enough energy. The terms come from the names of Enrico Fermi[1] and Satyendranath Bose.[2] The distinction between Bosons and Fermions is relevant to angular momentum, that is, to the way the wave function of a single particle transforms under rotation. The Pauli principle says that the permutation symmetry of a many-particle wave function (also called the *statistics* of the particle) depends on its spin. Why this connection? This is a new law of Nature that cannot be deduced from the postulates of Quantum Mechanics alone. However, in relativistic Quantum Mechanics the wave equations depend on spin: one has the Klein–Gordon equation for spin 0, the Dirac equation for spin 1/2, for the massive particles of spin 1, the Proca equation and so on. All those equations are untenable as single-particle equations, because of negative-energy solutions and other severe reasons. They can only stand as field equations, and the quanta of these fields are the particles. Then, the requirement that the field have a ground state (the vacuum state) is mandatory, and Fermion fields have one only if their wave functions are antisymmetric, while Bose fields require total permutation symmetry. In *relativistic field theory*, this is the spin-statistics theorem (Pauli 1940), but in the non-relativistic theory, it must be taken as a new principle. The implications of the Pauli principle range from the structure of atoms to superfluidity to the cohesion of solids and the stability of neutron stars. This topic has many aspects. In this chapter, we shall see how we can build a wave

[1] Enrico Fermi (Rome 1901–Chicago 1954) was one of the great scientists of the twentieth century, both for theoretical and experimental achievements. Among his contributions: the Fermi statistics, the theory of the Weak Interactions and the discovery of the Δ^{++} resonance. He was awarded the Nobel price in 1938 and he took advantage of the trip to Stockholm to quit the then fascist Italy for America, where he became a citizen and built the first nuclear reactor.

[2] Satyendranath Bose, Indian physicist (Calcutta 1894–1974), developed the statistics of integer spin particles. His work was completed in cooperation with Albert Einstein.

function having the correct permutation symmetry for a system of particles in terms of the single-particle wave functions, or spin-orbitals $\psi(\vec{r})$. In Chapter 21, we shall meet approximate methods (Hartree–Fock approximation and Density Functional) aimed at the best spin-orbitals $\psi(\vec{r})$.

19.2 Two Identical Particles

19.2.1 Fermions

Using two orthogonal spin-orbitals $\psi_1(\boldsymbol{x}_1)$, $\psi_2(\boldsymbol{x}_2)$, one must write a two-electron wave function in agreement with the Pauli principle. The antisymmetric wave function, or *Slater determinant,* setting a Fermion in each of the two spinors reads as

$$\begin{aligned}\Psi_{ab}(1,2) &= \frac{1}{\sqrt{2}}\left[\psi_a(1)\psi_b(2) - \psi_a(2)\psi_b(1)\right] \\ &= \frac{1}{\sqrt{2!}}\begin{vmatrix}\psi_a(1) & \psi_b(2)\\ \psi_a(1) & \psi_b(2)\end{vmatrix} \equiv |\psi_a\psi_b|. \end{aligned} \tag{19.3}$$

$\Psi(1,2)$ vanishes if $\psi_1 = \psi_2$; one cannot put two electrons in the same spin-orbital. This rule is the Pauli principle. N electrons occupy N distinct spin-orbitals. In the ground state, the lowest N spin-orbitals are occupied (the *Aufbau* principle). This is the ground configuration of atoms and molecules.[3]

Since many one-body states are degenerate, the ground state configuration built in this way is not unique in general. A common misconception leads some people to *believe* in the existence of individual spin-orbitals where the electrons belong in the many electron states. Equation (19.3) appears to say that the two electrons are in single-particle states ψ_a and ψ_b and the two-particle state is the Slater determinant. However, any unitary transformation, like $\psi_a = \alpha\xi_a + \beta\xi_b$, $\psi_b = -\beta^*\xi_a + \alpha^*\xi_b$ with $|\alpha|^2 + |\beta|^2 = 1$, gives[4] the same determinant, hence the same state, but with different orbitals. While the Slater determinant represents a possible approximate wave function of a physical two-particle state, there is no way to cage individual electrons into unique spin-orbitals. This remark extends to any many-body state and is one of the manifestations of entanglement (see Chap. 27).

[3] This simple rule holds as long as we can neglect the Coulomb interactions, which can mix different configurations. The antisymmetry principle always holds.

[4] As a consequence of $det(XY) = det(X)det(Y)$.

19.2.2 Bosons. Second Quantization

For N identical Bosons, on the other hand, one must symmetrize. Instead of determinants one writes *permanents* (all the minus signs become plus). Let $\phi_1(\boldsymbol{x}_1), \phi_2(\boldsymbol{x}_2)$ denote two wave functions for a boson field. They could represent wave functions of α particles or the wave vectors of two normal modes of the electromagnetic field. A two-boson wave function having the correct permutation symmetry can be taken in the form:

$$\Phi_{ab}(1,2) = \frac{1}{\sqrt{2}}\left[\phi_a(1)\phi_b(2) + \phi_a(2)\phi_b(1)\right], \tag{19.4}$$

where 1 and 2 stand for labels for the single-Boson states. One can represent the same two-body state in second quantization in the following form:

$$|ab\rangle = a_1^\dagger a_2^\dagger |\text{vac}\rangle. \tag{19.5}$$

Here, $a_1^\dagger, a_2^\dagger$ are the creation operators of the harmonic oscillators (known from Chap. 16) whose quanta have the space–time dependence $\phi_a(\boldsymbol{x}, t)$ and $\phi_b(\boldsymbol{x}, t)$), while $|\text{vac}\rangle$ stands for the mathematical vacuum, i.e., the absence of all the oscillators (actually in the physical vacuum all the oscillators compatible with the boundary conditions exist but are in their ground states). Thus, Eqs. (19.4) and (19.5) contain the same information, and the quantum states built through the two methods are in complete one-to-one correspondence if one takes that all the operators of different modes commute ($[a_1^\dagger, a_2^\dagger]_- = 0$, $[a_1^\dagger, a_2]_- = 0$ and the like). The second quantization, invented by Dirac, is a clever notation, and in addition, it is designed to handle the transitions between states with different numbers of bosons more easily. For each mode, the known commutation rules $[a, a^\dagger]_- = 1$ apply.

In summary, the Pauli theorem says that for two particles, the admissible wave functions built from one-particle wave functions are

$$\begin{cases} \psi_a(1)\psi_b(2) & \text{different particles,} \\ \frac{1}{\sqrt{2}}\left[\phi_a(1)\phi_b(2) - \phi_a(2)\phi_b(1)\right] & \text{identical fermions,} \\ \frac{1}{\sqrt{2}}\left[\phi_a(1)\phi_b(2) + \phi_a(2)\phi_b(1)\right] & \text{identical bosons.} \end{cases} \tag{19.6}$$

However, in general, the two-particle wave functions cannot be written down so simply in terms of independent particle wave functions.

19.2.3 Helium

Here I wish to introduce the simplest atom (beyond Hydrogen) and the simplest approximation to a two-electron (and many-electron) wave function, namely, the determinantal wave function. In this section, we are interested in the qualitative

aspects of the atomic wave function, and mainly on the symmetry; the actual numbers can be obtained by the variational method of Chap. 21, but are not that interesting (quantitative evaluations require more advanced techniques, anyhow).

The Hamiltonian is, in obvious notation,

$$H = \sum_{i=1}^{2} \left[\frac{p_i^2}{2m} - \frac{2e^2}{r_i} \right] + \frac{e^2}{r_{12}}. \tag{19.7}$$

Since H commutes (trivially) with the electron spin operators, the energy eigenstates are also spin eigenstates. Therefore, we can label all the energy eigenstates by the eigenvalues S, m_S of the total spin S^2, S_z. By the Pauli principle, since χ_S is odd under $P(1,2)$, the singlet orbital functions are even under the exchange of $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ and the triplet wave function are odd. Moreover, H has no angular dependence, and the electron states are eigenstates of L^2 and L_z.

The ground state turns out to be singlet and with a symmetric orbital wave function.

According to the Pauli principle, we can *put*[5] the two electrons in the spin-orbital states $\psi_1 = a(\overrightarrow{r})\alpha, \quad \psi_2 = a(\overrightarrow{r})\beta$; a more expressive notation is

$$\psi_1 = a(\overrightarrow{r}) \uparrow, \quad \psi_2 = a(\overrightarrow{r}) \downarrow . \tag{19.8}$$

The symmetry allows to write the ground state as

$$\Phi(1,2) = |a(\overrightarrow{r}) \uparrow a(\overrightarrow{r}) \downarrow |, \tag{19.9}$$

that is,

$$\begin{aligned} \Phi(1,2) &= \tfrac{1}{\sqrt{2}} [\psi_1(1)\psi_2(2) - \psi_1(2)\psi_2(1)] \\ &= a(\overrightarrow{r}_1)a(\overrightarrow{r}_2) \tfrac{[\alpha(1)\beta(2) - \alpha(2)\beta_2(1)]}{\sqrt{2}} = a(\overrightarrow{r}_1)a(\overrightarrow{r}_2)\chi_S. \end{aligned} \tag{19.10}$$

Here,

$$\chi_S = \frac{\alpha(1)\beta(2) - \beta(1)\alpha(2)}{\sqrt{2}} \tag{19.11}$$

is the spin singlet, such that

$$S^2\chi_S \equiv (S(1) + S(2))^2\chi_S = 0.$$

The assumption of a determinantal wave function means that we are assigning an electronic configuration to Helium, namely, a $1s^2$ ground configuration, with two opposite spin electrons in a 1 s level with a modified nuclear charge Z. In this scheme, Li has $1s^2 2s$, Be has $1s^2 2s^2$, B $1s^2 2s^2 2p$ and so on. The configurations become

[5] The Hartree–Fock method, to be discussed in Chap. 21, seeks the optimum spin-orbitals; the hydrogenic orbitals for Z = 2 are adequate for the present discussion.

gradually less useful with increasing the atomic number Z of the element because different configurations become closer in energy and tend to mix.

The triplet ($S = 1$) combinations are even:

$$\chi^T_{M_S} = \begin{cases} \chi^T_{M_S=1} = \alpha(1)\alpha(2), \\ \chi^T_{M_S=0} = \frac{\alpha(1)\beta(2)+\beta(1)\alpha(2)}{\sqrt{2}}, \\ \chi^T_{M_S=-1} = \beta(1)\beta(2). \end{cases} \tag{19.12}$$

They admit odd orbital functions. Letting b denote the 2s orbital of He, one can model the excited states of the $1s2s$ configuration:

$$\begin{array}{l} D_{\uparrow\uparrow} = |a \uparrow b \uparrow|, \; D_{\uparrow\downarrow} = |a \uparrow b \downarrow|, \\ D_{\downarrow\uparrow} = |a \downarrow b \uparrow|, \; D_{\downarrow\downarrow} = |a \downarrow b \downarrow|. \end{array} \tag{19.13}$$

For example, with the notation $a(1) \equiv a(\overrightarrow{r}_1)$, $b(1) \equiv b(\overrightarrow{r}_1)$ and so on,

$$D_{\uparrow\downarrow} = a(1)b(2)\alpha(1)\beta(2) - a(2)b(1)\alpha(2)\beta(1)$$

is correctly antisymmetric, but it is no singlet and no triplet. Since H is spin-independent, we have want spin eigenstates. Indeed,

$$\Psi_{S=0} = \frac{D_{\uparrow\downarrow} + D_{\downarrow\uparrow}}{\sqrt{2}} = \frac{a(1)b(2) + b(1)a(2)}{\sqrt{2}} \chi^S \tag{19.14}$$

and

$$\Psi_{S=1} = \frac{a(1)b(2) - b(1)a(2)}{\sqrt{2}} \chi^T_{M_S}; \tag{19.15}$$

$M_S = 1$ is just $D_{\uparrow\uparrow}$, $M_S = -1$ is $D_{\downarrow\downarrow}$. These are single Slater determinants; for $M_S = 0$, one needs the combination $\frac{D_{\uparrow\downarrow}+D_{\downarrow\uparrow}}{\sqrt{2}}$.

The Exchange Interaction

Let us compute the expectation value of the Coulomb interaction $\frac{e^2}{r_{12}} \equiv \frac{e^2}{|\boldsymbol{r}_1-\boldsymbol{r}_2|}$ over the two-electron eigenstates of spin. The spin averages yield 1.

For the triplets, the results are independent of M_s. One finds that

$$\begin{aligned} \langle\Psi_{\uparrow\uparrow}|\tfrac{e^2}{r_{12}}|\Psi_{\uparrow\uparrow}\rangle &= \tfrac{1}{2}\langle a(1)b(2) - b(1)a(2)|\tfrac{1}{r_{12}}|a(1)b(2) - b(1)a(2)\rangle \\ &= \tfrac{1}{2}[\langle a(1)b(2)|\tfrac{1}{r_{12}}|a(1)b(2)\rangle - \langle b(1)a(2)|\tfrac{1}{r_{12}}|a(1)b(2) \\ &-\langle a(1)b(2)|\tfrac{1}{r_{12}}|b(1)a(2)\rangle + \langle b(1)a(2)|\tfrac{1}{r_{12}}|b(1)a(2)\rangle]; \end{aligned} \tag{19.16}$$

since 1 and 2 are dummy indices,

$$\left\langle \Psi_{\uparrow\uparrow} | \frac{e^2}{r_{12}} | \Psi_{\uparrow\uparrow} \right\rangle = (a(1)b(2) - b(1)a(2)) | \frac{1}{r_{12}} | a(1)b(2)). \tag{19.17}$$

The first term

$$C = \langle a(1)b(2) | \frac{1}{r_{12}} | a(1)b(2) \rangle = \int d^3 r_1 d^3 r_2 \frac{|a(r_1)|^2 |b(r_2)|^2}{r_{12}} \tag{19.18}$$

has a clear electrostatic meaning and is called the *Coulomb term*; the second contribution

$$-E_x = -(b(1)a(2) | \frac{1}{r_{12}} | a(1)b(2))$$

is the *exchange term*. E_x is always positive, so its contribution is negative for the triplet; it turns out to be positive for the singlet $\Psi_{S=0}$.

Thus, the metastable triplet $1s2s^3S$ of He is below the singlet $1s2s^1S$ by an amount $2E_x$. The exchange term is always attractive for reasons that can be easily understood. The triplet wave function is antisymmetric when one exchanges $\overrightarrow{r}_1$ with $\overrightarrow{r}_2$ and the amplitude gets small at small distances. In other terms, one can imagine each electron that moves surrounded by the so-called *Fermi hole* or *exchange hole*, which is a region where it is unlikely to encounter other electrons of its own spin.[6] By this simple mechanism, high spin leads always to lower repulsion. This favors parallel spins. According to Hund's first rule, the ground state of any atom has the maximum possible spin compatible with its electronic configuration. This is the basic reason for the existence of ferromagnetism (and other forms of magnetic order) in some solids. The exchange term plays a key role in explaining the cohesion of solids and the covalent bond in general.

The above is just a rough introduction to some qualitative features of Atomic Physics. In atoms and in condensed matter, the electrons move in the potential of the nuclei and the collective potential due to the electron–electron interactions, besides the external fields used by the experiment. Since the measurements can be very accurate, the theory must take into account several effects. The spin–orbit interaction is proportional to $\overrightarrow{l} \cdot \overrightarrow{s}$ for each electron. Moreover, the electrostatic Coulomb interaction is only a first approximation to the electron–electron interactions, because of course the electrons move. The lowest-order relativistic corrections to the Coulomb interaction can be calculated in the *Breit approximation* that includes spin–spin, orbit–orbit and spin–other orbit interactions. Today, the atomic calculations are fully relativistic and make use of configuration mixing.

[6] This is not the whole story, because in addition to the exchange hole, which is explained by the Hartree–Fock approximation, that we shall present in Chap. 21, there is also a correlation hole where it is less likely to encounter other electrons of any spin; however, this is not understood by tha Hartree–Fock approximation and requires a much more accurate analysis.

19.3 Many-Body Matrix Elements. Fermions

For N independent electrons, the Hamiltonian is just $H = \sum_i^N h(i)$, where h is a single-particle operator, and Schrödinger's equation $H\Psi = E\Psi$ is trivially separable. One can solve in terms of the single-particle orthonormal spinor solutions of $h(i)\psi_\mu(i) = \epsilon_\mu \psi_\mu(i)$. Now, a product of one-electron solutions solves the Schrödinger equation, but cannot be accepted because of the Pauli principle. Never mind, we solve by setting

$$\Psi(1, 2, \ldots, N) = \hat{A} \prod_i^N \psi_{\mu(i)}(i), \tag{19.19}$$

with $E = \sum_\mu \epsilon(\mu)$; here $\hat{A}$ is the antisymmetrizer that converts the product to a Slater determinant; accordingly, in order to normalize the wave function, one must multiply the N electron antisymmetrized product by $\frac{1}{\sqrt{N!}}$. All the N electron states must be normalized linear combinations of Slater determinants made with spin-orbitals.

Now consider the set of orthonormal spin-orbitals that are eigenspinors of some one-body Hamiltonian h such that $h\psi_\mu = \epsilon_\mu \psi_\mu$. For three non-interacting electrons, a convenient concise notation for a normalized eigenfunction of the total Hamiltonian $H = h(a) + h(b) + h(c)$ with energy $E_{abc} = \epsilon_a + \epsilon_b + \epsilon_c$ is of the form

$$\Phi_{a,b,c}(1, 2, 3) = |\psi_a \psi_b \psi_c| = \frac{1}{\sqrt{3!}} det \begin{pmatrix} \psi_a(1) & \psi_a(2) & \psi_a(3) \\ \psi_b(1) & \psi_b(2) & \psi_b(3) \\ \psi_c(1) & \psi_c(2) & \psi_c(3) \end{pmatrix}. \tag{19.20}$$

This means $\Phi_{a,b,c}(1, 2, 3)\sqrt{3!} = \psi_a(1)\psi_b(2)\psi_c(3) + \psi_a(2)\psi_b(3)\psi_c(1) + \psi_a(3)\psi_b(1)\psi_c(2) - \psi_a(1)\psi_b(3)\psi_c(2) - \psi_a(3)\psi_b(2)\psi_c(1) - \psi_a(2)\psi_b(1)\psi_c(3)$. One-body operators depend on the coordinates (and spin) of one electron at a time, that is, are of the form $\hat{F} = \hat{f}(1) + \hat{f}(2) + \hat{f}(3)$. One can easily verify that the expectation values are

$$\langle \Phi_{a,b,c} | \hat{F} | \Phi_{a,b,c} \rangle = \langle \psi_a | \hat{f} | \psi_a \rangle + \langle \psi_b | \hat{f} | \psi_b \rangle + \langle \psi_c | \hat{f} | \psi_c \rangle. \tag{19.21}$$

Thus, the expectation values of momentum, angular momentum, kinetic energy and the like are the sum of those of the spin-orbitals. The charge density is

$$\langle \Phi_{a,b,c} | \hat{\rho}(\boldsymbol{x}, t) | \Phi_{a,b,c} \rangle = |\psi_a(\boldsymbol{x}, t)|^2 + |\psi_b(\boldsymbol{x}, t)|^2| + \psi_c(\boldsymbol{x}, t)|^2. \tag{19.22}$$

Here, $|.|^2$ denotes the square norm of spinors. One can show that the same rule applies for any number N of electrons.

One can verify easily for three electrons and also prove in general for any number of electrons that all the out-of-diagonal elements between determinants of one-body operators vanish if the states differ by two or more spin-orbitals. If the states differ

by one spin-orbital, one obtains the matrix element between the different orbitals. For instance, for three electrons,

$$\langle \Phi_{abc} | \boldsymbol{p}(1) + \boldsymbol{p}(2) + \boldsymbol{p}(3) | \Phi_{dbc} \rangle = \langle \psi_a(1) | \boldsymbol{p}(1) | \psi_d(1) \rangle. \tag{19.23}$$

For a prototype two-body operator, I introduce the Coulomb interaction, which I write in obvious notation:

$$H_C = \sum_{i<j} \frac{1}{r_{ij}} = \frac{1}{r_{12}} + \frac{1}{r_{13}} + \frac{1}{r_{23}}. \tag{19.24}$$

Developing the expectation value $\langle \Phi_{abc} | H_C | \Phi_{abc} \rangle$ with the determinant (19.20), one gets $6*3*6 = 108$ terms.

So, I introduce a lighter notation, writing $a(1)$ instead of $\psi_a(1)$ and the like. Then I observe that the six terms arising from the development of left determinant are permutations of 1, 2 and 3. We can apply the inverse permutation to all the terms. From odd permutations we collect a minus sign from both determinants. So one can write

$$\langle \Phi_{abc} | H_C | \Phi_{abc} \rangle =$$
$$\left\langle a(1)b(2)c(3) \left| \frac{1}{r_{12}} + \frac{1}{r_{13}} + \frac{1}{r_{23}} \right| \det \begin{pmatrix} a(1) & a(2) & a(3) \\ b(1) & b(2) & b(3) \\ c(1) & c(2) & c(3) \end{pmatrix} \right\rangle, \tag{19.25}$$

and these are 18 terms. Consider the terms in $\frac{1}{r_{12}}$. Since there is no operator acting on 3, only the terms with $c(3)$ multiplied by $c(3)$ survive. Therefore, the contribution is

$$\left\langle a(1)b(2) | \frac{1}{r_{12}} | \begin{pmatrix} a(1) & b(1) \\ a(2) & b(2) \end{pmatrix} \right\rangle.$$

Hence, the result is

$$\langle \Phi_{abc} | H_C | \Phi_{abc} \rangle = \left\langle a(1)b(2) \left| \tfrac{1}{r_{12}} \right| \begin{pmatrix} a(1) & b(1) \\ a(2) & b(2) \end{pmatrix} \right\rangle + \langle a(1)c(3) \left| \tfrac{1}{r_{13}} \right| \begin{pmatrix} a(1) & c(1) \\ a(2) & c(2) \end{pmatrix} \Big\rangle$$
$$+ \langle b(2)c(3) \left| \tfrac{1}{r_{23}} \right| \begin{pmatrix} b(2) & c(2) \\ b(3) & c(3) \end{pmatrix} \Big\rangle.$$

This can be developed and simplified to read as

$$\langle \Phi_{abc} | H_C | \Phi_{abc} \rangle = C - E, \tag{19.26}$$

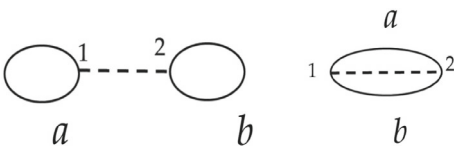


Fig. 19.1 Graphical representation of the direct term (19.27) (left) and exchange term (19.28); the sum over pairs of spin-orbitals is implied

where the Coulomb term is

$$C = \langle a(1)b(2)\left|\frac{1}{r_{12}}\right|a(1)b(2)\rangle + \langle a(1)c(3)\left|\frac{1}{r_{13}}\right|a(1)c(3))\rangle + \langle b(2)c(3)\left|\frac{1}{r_{23}}\right|b(2)c(3)\rangle; \tag{19.27}$$

the exchange term is

$$E = \langle a(1)b(2)\left|\frac{1}{r_{12}}\right|a(2)b(1)\rangle + \langle a(1)c(3)\left|\frac{1}{r_{13}}\right|a(3)c(1))\rangle + \langle b(2)c(3)\left|\frac{1}{r_{23}}\right|b(3)c(2)\rangle; \tag{19.28}$$

and it vanishes for antiparallel spins, since the matrix elements imply spin scalar products. The treatment extends directly to a many-body system where the sums extend to all pairs (Fig. 19.1).

Next, we consider the off-diagonal terms in a many-body system. If just one spin-orbital differs from bra and ket, the treatment is similar. Suppose there is a different spin-orbital d in the ket. Then, we must consider

$$\langle \Phi_{abc}|H_C|\Phi_{abd}\rangle = \left\langle a(1)b(2)c(3)\left|\frac{1}{r_{12}}+\frac{1}{r_{13}}+\frac{1}{r_3}\right|det\begin{pmatrix} a(1) & a(2) & a(3) \\ b(1) & b(2) & b(3) \\ d(1) & d(2) & d(3)\end{pmatrix}\right\rangle. \tag{19.29}$$

The result vanishes if the spin of d is opposite to the spin of c, since H_C commutes with the z component of the total spin. Therefore, assume that the spins of c and d are the same, but $\langle c|d\rangle = 0$. The terms in $\frac{1}{r_{12}}$ vanish, but $\frac{1}{r_{13}}$ and $\frac{1}{r_{23}}$ do contribute. The result is

$$\langle \Phi_{abc}|H_C|\Phi_{abd}\rangle = \sum_k^{occ}\left[\langle k(1)c(2)|\frac{1}{r_{12}}|k(1)d(2)\rangle - \langle k(1)c(2)|\frac{1}{r_{12}}|k(1)d(2)\rangle\right], \tag{19.30}$$

where $\sum_k^{occ}$ sums over the occupied states and the second term is the exchange contribution (for parallel spins) (Fig. 19.2). If both spin-orbitals are different in the ket, the calculation is easy:

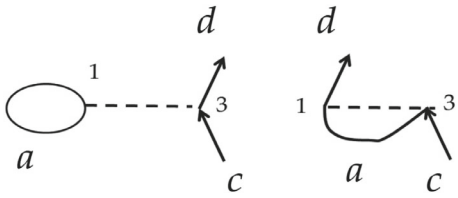


Fig. 19.2 Graphical representation of the direct (left) and exchange terms in the Coulomb scattering $ab \to ac$

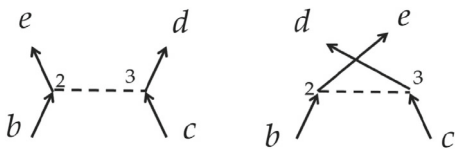


Fig. 19.3 Graphical representation of the direct (left) and exchange terms in the Coulomb scattering $ab \to de$

$$\langle \Phi_{abc}|H_C|\Phi_{aed}\rangle =$$
$$\left\langle a(1)b(2)c(3)\left|\frac{1}{r_{12}}+\frac{1}{r_{13}}+\frac{1}{r_3}\right|\det\begin{pmatrix} a(1) & a(2) & a(3)\\ e(1) & e(2) & e(3)\\ d(1) & d(2) & d(3)\end{pmatrix}\right\rangle, \tag{19.31}$$

but $\frac{1}{r_{12}}$ and $\frac{1}{r_{13}}$ yield nothing by orthogonality, and one is left with

$$\langle \Phi_{abc}|H_C|\Phi_{aed}\rangle = \langle b(2)c(3)\left|\frac{1}{r_{23}}\right|e(2)d(3)\rangle - \langle b(2)c(3)\left|\frac{1}{r}\right|d(2)e(3)\rangle, \tag{19.32}$$

where the second term is of the exchange type (Fig. 19.3).

As a result of a rather involved calculation, we obtained a simple rule that extends directly to any number of electrons.

19.3.1 Second Quantization for Fermions

In Sect. 16.2, the creator and annihilation operators for Boson excitations of a one-dimensional field, namely the harmonic oscillator, were presented. In a similar way, the second quantization formalism for Fermions allows us to deal with phenomena where particles are created or destroyed. One such phenomenon is the decay of the neutron into a proton, an electron and an antineutrino

$$n \to p + e + \bar{\nu};$$

in pair creation, an electron and a positron are created by γ rays, and many more reactions are known in elementary particle physics. In condensed matter physics, at much lower energies, the formalism is still useful. The promotion of an electron from a filled to an empty level can be described as the creation of an electron–hole pair.

In scattering processes, when all the particles are conserved, one can proceed with Slater determinants in the first quantization; however, second quantization formalism is generally preferred, since it is much more agreeable to work with; the picture is that the incoming particle is destroyed and the outgoing electron is created.

In order to change the formalism from Bosons to Fermions, we must simply replace permanents with determinants. In place of an N-times excited oscillator representing N bosons in a given mode, we now consider N-fermion determinants $|u_1 u_2 \dots u_N|$, where the spin-orbitals are chosen from a *complete orthonormal* set $\{w_i\}$. The index i can be discrete or continuous, but implies a *fixed ordering* of the complete set. In this way, one can convene, e.g., that in $|u_1 u_2 \dots u_N|$, the indices $1 \dots N$ are in increasing order, thereby avoiding multiple counting of the same state. The zero-particle or *vacuum* state $|\text{vac}\rangle$ replaces the oscillator ground state. For the determinants, it is generally preferable to use a compact notation like $|u_m u_n|$, rather than the explicit $\frac{1}{\sqrt{2}} Det \begin{pmatrix} u_m(1) & u_m(2) \\ u_n(1) & u_m(2), \end{pmatrix}$ which contains the same information. In second quantization, we aim for a unified description of all the Hilbert spaces with any number of particles, starting with the no-particle state that is the vacuum or $|vac\rangle$; the one-body state with one particle in spin-orbital u_k is denoted by $c_k^\dagger|vac\rangle$ and is described as the effect of the creation operator $c_k^\dagger$ over the vacuum. Similarly, the two-body determinantal state $|u_m u_n|$ becomes $c_m^\dagger c_n^\dagger|vac\rangle$, and the three-body determinant $|u_m u_n u_p|$ is written as $c_m^\dagger c_n^\dagger c_p^\dagger|vac\rangle$.

Since a determinant is odd when columns are exchanged, we want an anticommutation rule

$$[c_m^\dagger, c_n^\dagger]_+ \equiv c_m^\dagger c_n^\dagger + c_n^\dagger c_m^\dagger = 0. \tag{19.33}$$

It follows that the square of a creation operator vanishes.

Adding a particle to any state cannot lead to the vacuum state,

$$\langle vac|c_m^\dagger = 0. \tag{19.34}$$

By definition,

$$c_m^\dagger\{c_n^\dagger c_r^\dagger|vac\rangle\} = c_m^\dagger c_n^\dagger c_r^\dagger|vac\rangle. \tag{19.35}$$

The notation suggests that $c_m^\dagger$ is the Hermitian conjugate of c_m; this is called an annihilation operator. Taking the conjugate of (19.35)

$$\{\langle vac|c_r c_n\}c_m = \langle vac|c_r c_n c_m \tag{19.36}$$

and taking the scalar product with $c_m^\dagger c_n^\dagger c_r^\dagger|vac\rangle$, we deduce that

$$\{\langle vac|c_r c_n\}c_m \mid c_m^\dagger c_n^\dagger c_r^\dagger|vac\rangle = 1. \tag{19.37}$$

If we now consider c_m as acting on the right, we see that it is changing the three-body state $c_m^\dagger c_n^\dagger c_r^\dagger|vac\rangle$ into the two-body one $c_n^\dagger c_r^\dagger|vac\rangle$. Thus, annihilation operator is a

well-deserved name: an annihilation operator c_m for a fermion in the spin-orbital state u_m removes the leftmost state in the determinant, leaving a $N-1$ state determinant:

$$c_1|u_1u_2\ldots u_N| = |u_2\ldots u_N| \tag{19.38}$$

and

$$c_m|vac\rangle = 0. \tag{19.39}$$

It obeys the conjugate of the anticommutation rules (19.33), namely,

$$[c_m, c_n]_+ \equiv c_mc_n + c_nc_m = 0,\ c_m^2 = 0. \tag{19.40}$$

Next, consider

$$c_nc_m^\dagger c_n^\dagger c_r^\dagger|vac\rangle,\ n, m, r \ \text{ all different}. \tag{19.41}$$

Since the creation operators anticommute, we get

$$-c_nc_n^\dagger c_m^\dagger c_r^\dagger|vac\rangle = c_m^\dagger c_r^\dagger|vac\rangle,$$

since the n state is created at the leftmost place in the determinant, but is annihilated at once. This shows that creation and annihilation operators also anticommute,

$$[c_n, c_m^\dagger]_+ = 0,\ n \neq m. \tag{19.42}$$

As long as the indices are different, c and $c^\dagger$ all anticommute, so the pairs c_nc_m,$c_nc_m^\dagger$,$c_n^\dagger c_m$ and $c_n^\dagger c_m^\dagger$ can be carried through any product of creation or annihilation operators in which the indices n, m do not occur.

Next, we note that $c_p^\dagger|vac\rangle \equiv |p\rangle$ is a one-body wave function; $c_pc_p^\dagger|vac\rangle = |vac\rangle$ and $c_p^\dagger c_pc_p^\dagger|vac\rangle = c_p^\dagger|vac\rangle$. Now one can check that

$$n_p \equiv c_p^\dagger c_p \tag{19.43}$$

is the occupation number operator, since it has eigenvalue 1 on any determinant where p is occupied and 0 if p is empty. On the other hand, $c_pc_p^\dagger$ has eigenvalue 0 on any determinant where p is occupied and 1 if p is empty. Thus in any case $c_pc_p^\dagger + c_p^\dagger c_p = 1$. Since this holds on all the complete set it is an operator identity and we may complete the rules with

$$[c_p, c_q^\dagger]_+ = \delta_{pq}. \tag{19.44}$$

Note that $n_p^\dagger = n_p$ and $n_p^2 = n_p$.

It is important to be able to change basis, i.e., to switch from basis set $\{a_n\}$ to a new set $\{b_n\}$; since

$$|b_n >= \sum_k |a_k >< a_k|b_n >, \tag{19.45}$$

the rule is

$$b_n^\dagger = \sum_k a_k^\dagger < a_k|b_n >, \quad b_n = \sum_k a_k < b_n|a_k > . \tag{19.46}$$

It is often useful to go from any set $\{u_n\}$ to the coordinate representation introducing the creation and annihilation field operators

$$\begin{cases} \Psi^\dagger(\boldsymbol{x}) = \sum_n c_n^\dagger u_n^\dagger(\boldsymbol{x}) \\ \Psi(\boldsymbol{x}) = \sum_n c_n u_n(\boldsymbol{x}), \end{cases} \tag{19.47}$$

(here, $u_n^\dagger$ denotes the conjugate spinor). Note that $c_p^\dagger|vac\rangle$ is a one-electron state and corresponds to the first-quantized spinor $u_p(\boldsymbol{x})$; $\Psi^\dagger(\boldsymbol{y})|vac\rangle$ is a one-electron state and corresponds to the first-quantized spinor with spatial wave function $\sum_n u_n^\dagger(\boldsymbol{y})u_n(\boldsymbol{x}) = \delta(\boldsymbol{x} - \boldsymbol{y})$; thus, it is a perfectly localized electron. The rules are readily seen to be

$$[\Psi(\boldsymbol{x}), \Psi(\boldsymbol{y})]_+ = 0, \ [\Psi^\dagger(\boldsymbol{x}), \Psi^\dagger(\boldsymbol{y})]_+ = 0, \tag{19.48}$$

and

$$[\Psi^\dagger(\boldsymbol{y}), \Psi(\boldsymbol{x})]_+ = \sum_{p,q} [c\dagger_p, c_q]_+ u_p\dagger(\boldsymbol{x})u_q(\boldsymbol{y}) = \sum_{p,q} u_p\dagger(\boldsymbol{x})u_p(\boldsymbol{y}) = \delta(\boldsymbol{x} - \boldsymbol{y}), \tag{19.49}$$

where the δ also imposes the same spin for both spinors.

A one-body operator $V(x)$ in second-quantized form becomes

$$\hat{V} = \int dx \Psi^\dagger(x) V(x) \Psi(x) = \sum_{p,q} V_{p,q} c_p^\dagger c_q. \tag{19.50}$$

This gives the correct matrix elements between determinantal states, as one can verify.

The above expressions imply spin sum along with the space integrals, although this was not shown explicitly; let me write the spin components, for one-body operators:

$$\hat{V} = \sum_{\alpha,\beta} \int dx \Psi_\alpha^\dagger V_{\alpha,\beta}(x) \Psi_\beta. \tag{19.51}$$

For the spin operators, setting $\hbar = 1$, and using the Pauli matrices, $S_z = \frac{1}{2}\sigma_z$, $S^+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, and the rule (19.50), one finds

$$S_z = \frac{1}{2}\int dx \left(\Psi_\uparrow^\dagger(x)\Psi_\uparrow(x) - \Psi_\downarrow^\dagger(x)\Psi_\downarrow(x)\right), \quad S^+ = \int dx \Psi_\uparrow^\dagger(x)\Psi_\downarrow(x). \tag{19.52}$$

When using a discrete basis and notation, we shall write

$$S^+ = \sum_k c_{k\uparrow}^\dagger c_{k\downarrow}, \tag{19.53}$$

which is obtained from (19.52) by taking a Fourier transform in discrete notation. A two-body operator $U(x, y)$ becomes

$$\hat{U} = \int dx \int dy \Psi^\dagger(x)\Psi^\dagger(y)U(x, y)\Psi(y)\Psi(x) = \sum_{ijkl} U_{ijkl} c_i^\dagger c_j^\dagger c_l c_k \tag{19.54}$$

(please note the order of indices carefully). The Hamiltonian for N interacting electrons in an external potential $\varphi(x)$ is the *true* many-body Hamiltonian in the non-relativistic limit that we shall often regard as the full many-body problem for which approximations must be sought. It may be written as

$$H(r_1, r_2, \ldots, r_N) = H_0(r_1, r_2, \ldots, r_N) + U(r_1, r_2, \ldots, r_N), \tag{19.55}$$

where H_0 is the free part

$$H_0 = T + V_{ext} = \sum_i \left\{-\frac{1}{2}\nabla_i^2 + V(r_i)\right\} = \sum_i h_0(i), \tag{19.56}$$

with T the kinetic energy and V_{ext} the external potential energy, while

$$U = \frac{1}{2}\sum_{i\neq j} u_C(r_i - r_i) \tag{19.57}$$

is the Coulomb interaction. This Hamiltonian may be written in second-quantized form

$$\begin{aligned} H &= H_0 + U, \\ H_0 &= \sum_\sigma \int dr \Psi_\sigma^\dagger(r) h_0 \Psi_\sigma(r), \\ U &= \frac{1}{2}\sum_{\alpha,\beta,\gamma,\delta} \int\int dx dy \psi_\alpha^\dagger(x)\psi_\beta^\dagger(y) u_C(x - y)_{\alpha\gamma,\beta\delta}\Psi_\delta(y)\Psi_\gamma(x). \end{aligned} \tag{19.58}$$

Often, the spin indices are understood as implicit in the integrations. It should be kept in mind that relativistic corrections are needed in most problems with light elements and the relativistic formulation is needed when heavy elements are involved. Fortunately, the ideas that we shall develop lend themselves to a direct generalization to Dirac's framework.

19.3.2 Fano–Anderson Models

As an example of the above formalism, consider the following first-quantized Hamiltonian H_{fq}, representing a discrete level $|0\rangle$ with energy ϵ_0, and a continuum of levels $|k\rangle$ with energy ϵ_k;

$$H_{fq} = \sum_k \epsilon_k |k\rangle\langle k| + \epsilon_0 |0\rangle\langle 0| + \sum_k [V_k |k\rangle\langle 0| + V_k^* |0\rangle\langle k|]. \tag{19.59}$$

The last term allows for a mixing between the continuum and the discrete level. This model represents a single particle (Fermi or Bose) that can hop from a localized state to a band of states and back. The second-quantized version of the model is the Fano model

$$H_F = \sum_{k,\sigma} \epsilon_k \hat{n}_{k\sigma} + \epsilon_0 \hat{n}_{0\sigma} + \sum_k [V_k c_{k\sigma}^\dagger c_{0\sigma} + h.c.]. \tag{19.60}$$

(Here, h.c. stands for the Hermitian conjugate of the preceding term.)

This represents the same physics as (19.59) if it is taken to act in single-particle subspace (but now the anticommuting creation operators are for spinors and σ denotes spin; the solution to this model is very instructive and is deferred to Chap. 24).

The same formalism lends itself to important extensions. It allows us to work with many-body systems where the electron states are filled up to a Fermi level. Finally, we may add an interaction term:

$$H_A = \sum_{k,\sigma} \epsilon_k \hat{n}_{k\sigma} + \epsilon_0 \hat{n}_{0\sigma} + \sum_k [V_k c_{k\sigma}^\dagger c_{0\sigma} + h.c.] + U n_{0+} n_{0-}. \tag{19.61}$$

This is called the Anderson impurity model and has been widely used to describe magnetic impurities in metals with many interesting many-body complications. We shall say more on it in Chap. 24.

19.4 Quasi-particles

19.4.1 Boson and Fermion Excitations

The quantum many-body problem in atoms and condensed matter systems is extremely involved, since the electron–electron interaction is always important, but some excitations allow a simple effective description reminiscent of non-interacting models. For instance, the excitation that corresponds to the threshold of light absorption in a semiconductor, in a rigorous treatment involves all the electrons and the nuclei of the system; yet, it admits a simple, rather accurate and enlightening description as a one-electron interband transition in an independent particle model.

However, in this description, the fermion is not simply the "bare" electron, but has a renormalized mass, due to interaction with the lattice, with the phonons and with the other electrons; moreover, it has a modified gyromagnetic ratio and a finite lifetime. In other words, it is a quasi-electron. The absence of an electron below the Fermi level[7] gives rise to complicated excited states of the many-body system that are well described by Fermion quasi-particles, the so-called holes. The low-lying excited states are well described by weakly interacting excitations called quasi-particles. Such a description, originally due to Lev Landau, is borne out by the modern many-body Green's function formalism and is extremely useful for describing complex situations quantitatively and figuring out what happens if we perturb the system. Even in strongly correlated systems, the low-energy excitations behave as Fermions or Boson weakly interacting particles, that are known as quasi-particles since they have a finite width in energy that can be interpreted as a finite lifetime. Many excitations are quantized boson quasi-particles, like the phonons, charge density waves or spin waves. Electron and hole quasi-particles can also combine to form *excitons* which are electron–hole bound states. Frenkel excitons are small radius excitons and dominate the optical properties of insulators and organic compounds, while Wannier–Mott excitons typical of GaAs and other semiconductors have binding energies of less than one eV.

Some of the higher energy excitations (*plasmons*, or quantized plasma oscillations) are also oscillator-like and can be treated as Bosons. Plasmons are evident in photoemission spectra of metals and semiconductors as satellite lines of all the prominent features. The plasmon frequency at long wavelengths is

$$\omega_p = \sqrt{\frac{4\pi n e^2}{m}}, \tag{19.62}$$

where n is the electron number density, e the electron charge and m the electron mass.

[7] The Fermi level is the top occupied energy level at absolute zero; Chap. 25 is devoted to Quantum Statistical Mechanics.

Classically, plasmons can be understood as being due to the collective oscillation of the electrons in a slab against a positive background. A shift η produces a surface charge density $\sigma = \eta ne$, and the resulting electric field $E = 4\pi\sigma$ orthogonal to the surface causes the oscillations across the slab; it is a simple exercise to derive Equation (19.62). The plasmons are the quanta of this oscillator. At solid surfaces one observes surface plasmons, in which all long wavelengths have frequency $\frac{\omega_p}{\sqrt{2}}$. One can show that in a spherical metal particle the basic dipole surface plasmon oscillation has a frequency $\frac{\omega_p}{\sqrt{3}}$. In a more accurate description one takes into account that the plasmons are not sharp because they can decay into electron–hole excitations.

A *polariton* at metal surfaces is mixed surface plasmon–photon excitation, which is predominantly photon-like at small wave vectors and becomes plasmon-like at short wavelengths. In superconductors, the current is carried by *Cooper pairs* (see Chap. 25). A quantitative description of quasi-particles is obtained through field theory methods, often built in computer codes. Nevertheless, the qualitative analysis of the problems remains important.

19.4.2 Anyons

In two-dimensional systems, quasi-particles can be not only Bosons or Fermions, but, more generally, Anyons. Excited states of systems exhibiting the Integer Quantum Hall effect (see Chap. 26) support anyons.[8] These are quasi-particles that are possible only in two-dimensional models. An exchange P_{12} of Anyons 1 and 2 multiplies the wave function of the system by some phase factor different from ± 1. Indeed, a double exchange P_{12}^2 is equivalent to a round trip of particle 1 around particle 2, which leaves the system in the same state. The same state means the same wave function, up to an arbitrary phase factor. We may write $P_{12} = e^{i\phi}$ with real ϕ. In three-dimensional models, one can deform any close path continuously to a point, and this implies that $\phi = 0 \Rightarrow P_{12}^2 = 1 \Rightarrow P_{12} = \pm 1$. The only possibilities are Bosons and Fermions, as we know. In two-dimensional models, however, one cannot reduce the path to a point without crossing particle 2. Therefore, any ϕ is possible.

One can obtain a semi-classical mental picture of an anyon as a spinless particle with charge q orbiting around a thin solenoid at distance r. If the current in the solenoid vanishes (i = 0), then the component L_z of the orbital angular momentum orthogonal to the orbit is obviously an integer. Now turn on the current i. The particle feels an electric field such that

$$\int rot\boldsymbol{E}.\boldsymbol{n}d^2r = -\frac{\partial}{\partial t}\int \boldsymbol{B}.\boldsymbol{n}d^2r = -\frac{\partial\phi}{\partial t}, \tag{19.63}$$

[8] See for instance the beautiful book by Jiannis K. Pachos, "Introduction to Topological Quantum Computation", Cambridge University Press.

where ϕ is the flux within the orbit. The e.m.f. is $E = -\frac{\frac{\partial \phi}{\partial t}}{2\pi r}$ and is directed in a plane, orthogonal to the radius. The electric field changes the angular momentum. The intensity of the torque is $rE = \frac{\frac{\partial \phi}{\partial t}}{2\pi} = \dot{L}_z$. Thus, the angular momentum is modified by $\Delta L_z = -\frac{q}{2\pi}\Delta\phi$.

In a more accurate description, the charge is being switched on with the flux, and the two are proportional. This leads to $\Delta L_z = -\frac{q}{4\pi}\Delta\phi$. The idea behind such formulations is that a time-dependent flux has a chirality and can impart angular momentum; the presence of the flux can be accepted since the anyon is actually a composite quasi-particle object.

Anyons do exist. They are excitations that form at very low temperatures in the electron liquid when a thin semiconductor layer is sandwiched between AlGaAs layers in strong magnetic field; they carry fractional units of magnetic flux in the Fractional Quantum Hall effect. Two anyons cannot occupy the same quantum state, but several of them can form time-dependent braids, in which the anyons enter with several possible topologies. The different braids contain information. It has been suggested by Alexei Kitaev in 1997 that the topological properties should be robust against the perturbations that produce decoherence (see Sect. 9.5.1), and therefore it should be possible to build logical gates and a quantum computer operating with anyons. Research on the topological quantum computer is ongoing.

19.5 Hellmann–Feynman Theorem and Virial Theorem

19.5.1 Hellmann–Feynman Theorem

One of the few exact results for interacting many-body problems runs as follows. Let $H = H(\lambda)$ denote a many-body Hamiltonian which depends on a parameter λ and E denote an eigenstate of H. Then $E(\lambda) = \langle \Psi | H(\lambda) | \Psi \rangle$ and a small $d\lambda$ produces a first-order variation $d\Psi$, which is orthogonal to Ψ. Therefore,

$$\frac{dE}{d\lambda} = \left\langle \Psi | \frac{dH}{d\lambda} | \Psi \right\rangle. \tag{19.64}$$

19.5.2 Virial Theorem

Consider a system of q particles (electrons and nuclei) in Coulomb interaction, as described by the Hamiltonian (in obvious notation)

$$H = -\frac{1}{2}\sum_{i=1}^{q} \frac{\nabla_i^2}{m_i} + \sum_{i<j}^{q} \frac{e_i e_j}{r_{ij}}. \tag{19.65}$$

The ground state wave function depends on the coordinates and spins of all the particles. We denote it $\Phi(x_{\alpha,i})$, with $\alpha = 1, 2, 3$ denoting the Cartesian component while i runs over the particles. Now consider the family of functions $\Phi(\eta x_{\alpha,i})$, where η is a scale factor. (Note that if $\eta = 2$, the wave function is contracted by a factor of 2.) Then we must normalize the wave function again and consider

$$\Phi_\eta = \Phi(\eta x_{\alpha,i})\eta^{3q/2}. \tag{19.66}$$

Let $\langle T\rangle^\eta$, $\langle H\rangle^\eta$ and $\langle V\rangle^\eta$ denote the new values calculated with the scaled wave functions. It is clear that

$$\langle H\rangle^\eta = \eta^2\langle T\rangle^{(1)} + \eta\langle V\rangle^{(1)}. \tag{19.67}$$

If we consider η as a variational parameter, we get the condition

$$2\eta\langle T\rangle^{(1)} + \langle V\rangle^{(1)} = 0. \tag{19.68}$$

So,

$$\eta = -\frac{\langle V\rangle^{(1)}}{2\langle T\rangle^{(1)}}. \tag{19.69}$$

But $\eta = 1$ is already the minimum, since we started from the exact ground state. This proves the virial theorem, which is one of the few known exact statements about interacting many-body systems.

19.6 Hint to Spontaneous Symmetry Breaking and Goldstone Theorem

Imagine a system that has a certain symmetry, geometrical or otherwise. This means that the laws of physics are the same when you perform a transformation on it (e.g., a rotation or a reflection). In some cases, the lowest energy state of the system doesn't respect this symmetry, even though the underlying laws do. Spontaneous symmetry breaking occurs when the lowest energy state of the system chooses a specific configuration that breaks the overall symmetry, even though the equations describing the system are symmetric. (This also applies to the ground state of a quantum system or to the vacuum state of a field, as discussed in the quantum part of this book.) This is a bit like a pencil balanced perfectly upright on its tip; it can fall in any direction, breaking the rotational symmetry.

The Goldstone theorem is a fundamental concept in quantum field theory that is related to spontaneous symmetry breaking. It was formulated by Jeffrey Goldstone in 1961 and plays a crucial role in understanding certain aspects of particle physics,

particularly in the context of the Higgs mechanism and the emergence of mass in gauge theories; although the first examples belong to condensed matter theory and are due to Phil Anderson.

The Goldstone theorem states that if a continuous symmetry is spontaneously broken in a quantum field theory, then there must exist massless scalar particles called Goldstone bosons. These bosons are associated with the broken symmetries and are responsible for the long-range forces associated with the broken symmetry.The Higgs mechanism in the Standard Model of particle physics is an example where the Goldstone theorem plays a role. The Higgs boson itself arises from the spontaneous breaking of electroweak symmetry and interacts with other particles to give them mass. However the concept is general and not limited to quantum field theory.

Chapter 20
Perturbation Theory

Qualitative and approximate methods are very important, since they allow us to understand trends. The simple techniques in this chapter help one's physical intuition very much.

20.1 Time-Independent Rayleigh–Schrödinger Perturbation Theory

Suppose we can find the bound states of some Hamiltonian H_0 by solving exactly the time-independent Schrödinger equation $H_0\psi_n^{(0)} = E_n^{(0)}\psi_n^{(0)}$. We calculate the discrete energy levels $\{E_1^{(0)}, E_2^{(0)}, E_3^{(0)}, \ldots\}$ and find agreement with the experiment. Invariably then, we wish to see what happens when we perturb the system by what we physically feel is a weak perturbation $\hat{V}$. In most cases, we discover that we are unable to solve

$$(H_0 + \hat{V})\psi_n^{(0)} = E_n^{(0)}\psi_n^{(0)}, \tag{20.1}$$

or perhaps we can solve, but the solution is too involved to make sense of, but if the perturbation is *weak,* E_n differing little from $E_n^{(0)}$, the changes are small compared to the gap between different unperturbed eigenvalues. Then, perturbation theory offers a simple approximate solution that is satisfactory for many purposes and allows us to grasp trends and make predictions. The idea is that of expanding in powers of the perturbation. In practice, nobody goes beyond second order or third order. With increasing order, the approach quickly becomes unwieldy, and one must look for a different route. Besides, in some important problems, the approach fails, and then it emerges that the solution is not an analytical function of V so the perturbation series is a bad start. If the spectrum is continuous, the perturbation is never small, and one

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_20

must resort to the field theory approaches which are a smarter form of perturbation theory taken to infinite order. On the other hand, the present approach can be useful in one-body and many-body problems, with or without spin, with Fermions or Bosons.

We expand the perturbed wave function of the level n over the unperturbed basis $\{\psi^{(0)}\}$:

$$\psi_n = \sum_m c_{n,m}\psi_m^{(0)}. \tag{20.2}$$

We write $\lambda\hat{V}$ for $\hat{V}$. At the end, $\lambda = 1$, but this trick helps to visualize the perturbation order of the various terms of the development. Then, from (20.1),

$$\sum_m c_{n,m}\left(E_m^0 + \lambda\hat{V}\right)\psi_m^{(0)} = E\sum_m c_{n,m}\psi_m^{(0)}. \tag{20.3}$$

Multiplying by ψ_k^{0*} and integrating, we obtain the exact algebraic restatement of the problem:

$$\boxed{(E_n - E_k^{(0)})c_{n,k} = \lambda\sum_m V_{km}c_{n,m}.} \tag{20.4}$$

Equation (20.1) is rewritten as a discrete infinite system that we are going to solve by successive approximations.

20.1.1 Discrete Non-degenerate Spectrum

In this case (different unperturbed eigenvalues $E_n^{(0)}$ for different n), the weak perturbation changes the levels so little (compared to the level separation) that one can tell, for each perturbed energy, the unperturbed one from which it is derived. Then, we look for the solution of (20.4) in the form of a series expansion that starts from $\psi_n^{(0)}$ for $\lambda\hat{V} \to 0$, assuming a series expansion in powers of λ,

$$\psi_n = \psi_n^{(0)} + \lambda\psi_n^{(1)} + \lambda^{(2)}\psi_n^{(2)} + \cdots. \tag{20.5}$$

Therefore, we set

$$\begin{aligned} c_{n,k} &= c_{n,k}^{(0)} + \lambda c_{n,k}^{(1)} + \lambda^{(2)}c_{n,k}^{(2)} + \cdots \\ E &= E_n^{(0)} + \lambda E_n^{(1)} + \lambda^{(2)}E_n^{(2)} + \cdots, \end{aligned}$$

with $c_{n,k}^{(0)} = \delta_{k,n}$. Then, (20.4) becomes

$$\begin{aligned} (E_n^{(0)} - E_k^{(0)} + \lambda E_n^{(1)} + \lambda^2 E_n^{(2)} + \cdots)(\delta_{k,n} + \lambda c_{n,k}^{(1)} + \lambda^2 c_{n,k}^{(2)} + \cdots) = \\ \lambda\ \textstyle\sum_m V_{km}(\delta_{m,n} + \lambda c_{n,m}^{(1)} + \lambda^2 c_{n,m}^{(2)} + \cdots). \end{aligned} \tag{20.6}$$

We multiply, and then we separate the different orders (the powers of λ):

$$\begin{aligned}[E_n^{(0)} - E_k^{(0)}]\delta_{k,n} &+ \lambda[E_n^{(1)}\delta_{k,n} + c_{n,k}^{(1)}(E_n^{(0)} - E_k^{(0)})] \\ &+ \lambda^2[E_n^{(2)}\delta_{k,n} + c_{n,k}^{(2)}(E_n^{(0)} - E_k^{(0)}) + c_{n,k}^{(1)}E_n^{(0)}] + \cdots \\ &= \lambda \sum_m V_{km}\delta_{m,n} + \lambda^2 \sum_m V_{km}c_{n,m}^{(1)} + \cdots .\end{aligned} \tag{20.7}$$

At order 0, we get the trivial result $(E_n^{(0)} - E_k^{(0)})\delta_{k,n} = 0$. At first order, however,

$$\underbrace{(E_n^{(0)} - E_k^{(0)})c_{n,k}^{(1)}} + \underbrace{E_n^{(1)}\delta_{k,n}} = \sum_m V_{k,m}\delta_{m,n} = V_{k,n}. \tag{20.8}$$

For $k = n$, the first curly bracket vanishes, and one finds

$$E_n^{(1)} = V_{n,n}. \tag{20.9}$$

For $k \neq n$, the second curly bracket vanishes, and one finds the amplitudes of the *virtual excitations* $n \leftrightarrows k$

$$c_{n,k}^{(1)} = \frac{V_{k,n}}{E_n^{(0)} - E_k^{(0)}}, \quad k \neq n,$$

and the first-order correction to $\psi_n^{(0)}$ is found to be

$$\psi_n^{(1)} = \sum_{k \neq n} \frac{V_{k,n}}{E_n^{(0)} - E_k^{(0)}}\psi_k^{(0)}. \tag{20.10}$$

The correction cannot blow up since the spectrum is discrete. Note that we need $< \psi^{(1)}|\psi^{(0)} >= 0$ in order to ensure that $\langle\psi|\psi\rangle = 1 + O(\lambda^2)$, that is, $\psi^{(1)}$ is correctly normalized (the error is at second order). It follows that $c_{n,n}^{(1)} = 0$.

From (20.10), it is clear that the first approximation is good if

$$\frac{V_{k,n}}{E_n^{(0)} - E_k^{(0)}} \ll 1.$$

In second order, (20.7) yields

$$\underbrace{(E_n^{(0)} - E_k^{(0)})c_{n,k}^{(2)}} + \underbrace{E_n^{(1)}c_{n,k}^{(1)} + E_n^{(2)}\delta_{k,n}} = \sum_m V_{k,m}c_{n,m}^{(1)}.$$

For $k = n$, the second curly bracket survives; since $c_{n,n}^{(1)} = 0$; the second-order correction to the energy reads as

$$E_n^{(2)} = \sum_m V_{n,m} c_{nm}^{(1)} = \sum_{m \neq n} \frac{|V_{n,m}|^2}{E_n^{(0)} - E_m^{(0)}}. \tag{20.11}$$

This is always negative for the ground state.

Problem 20.1 Show that the energy of the H atom in the ground state plunged into a weak uniform electric field $\mathbf{E}$ is $E = E_0 + \frac{1}{2}\alpha \mathbf{E}^2$ and show the order of magnitude of the polarizability α.

Solution 20.1 The effect on the ground state of the perturbation $\hat{V} = ez\mathbf{E}$ is quadratic, since the operator is odd and $\langle \psi_0 | \hat{V} | \psi_0 \rangle = 0$. The correction to the energy arises in second order. The polarizability has the dimensions of a volume and is about a_0^3. The experimental value is $\alpha = 0.66\ 10^{-24}\ \mathrm{cm}^3$.

Problem 20.2 The oscillator with Hamiltonian

$$H_0 = \frac{p^2}{2m} + \frac{1}{2} m\omega^2 x^2$$

is perturbed by

$$H' = bx + c.$$

Find the correction to the ground state energy by perturbation theory up to the second order in the case $b = -m\omega^2 a$ with $c = \frac{1}{2} m\omega^2 a^2$. Since this case corresponds to a shift $x \to x - a$, one can compare the perturbation theory approximation with the exact result.

Solution 20.2 Recall that $\hat{x} = \frac{x_0(a+a^\dagger)}{\sqrt{2}}$, $\hat{p} = -i\hbar \frac{(a-a^\dagger)}{x_0\sqrt{2}}$, where $x_0 = \sqrt{\frac{\hbar}{m\omega}}$, $[a, a^\dagger] = 1$. In first order, the average yields c, and in second order, $E^{(2)} = -\frac{b^2 (x_{01})^2}{\hbar\omega} = -\frac{b^2}{\hbar\omega}\frac{x_0^2}{2} = -\frac{1}{2}\frac{b^2}{m\omega^2}$. The correction up to second order gives

$$c - \frac{1}{2}\frac{b^2}{m\omega^2} = 0.$$

This is *exact*, since the effect of H' is $x \to x - a$. All the higher terms of the perturbation series sum up to 0.

Problem 20.3 For the *anharmonic* oscillator with Hamiltonian

$$\hat{H} = \left(\hat{n} + \frac{1}{2}\right)\hbar\omega + \alpha \left(\frac{x}{x_0}\right)^4,$$

calculate the effect of the perturbation on the ground state in first and second orders. When can we say that the second order is small compared to the first?

Solution 20.3 Recall that $\hat{x} = (x_0(a + a^\dagger))/\sqrt{2}$, $\hat{p} = -i\hbar((a - a^\dagger))/(x_0\sqrt{2})$, where $x_0 = \sqrt{\frac{\hbar}{m\omega}}$, $[a, a^\dagger] = 1$. The first-order correction is $E^1 = \alpha[(\frac{x}{x_0})^4]_{0,0}$. The nonzero matrix elements (16.28) of x are $x_{n+1,n} = x_{n,n+1} = C\sqrt{n+1}$, with $C = \frac{x_0}{\sqrt{2}}$. Acting on x, from $|0\rangle$, one arrives at $|1\rangle$, and so

$$(x^4)_{0,0} = x_{0,1}(x^2)_{1,1}x_{1,0}.$$

Therefore,

$$(x^2)_{1,1} = x_{1,0}x_{0,1} + x_{1,2}x_{2,1}.$$

So,

$$(x^4)_{0,0} = x_{0,1}[x_{1,0}x_{0,1} + x_{1,2}x_{2,1}]x_{1,0} = 3C^4,$$

and the first-order correction is $E^1 = \frac{3}{4}\alpha$.

In working out (20.11), the states with $m > 4$ play no role. Since by parity, $(x^4)_{0,1} = 0 = (x^4)_{0,3}$,

$$\begin{aligned}(x^4)_{0,2} &= x_{0,1}(x^3)_{1,2} = x_{0,1}[x_{1,2}(x^2)_{2,2} + x_{1,0}(x^2)_{0,2}]\\ &= x_{0,1}x_{1,2}(x_{2,1}x_{1,2} + x_{2,3}x_{3,2}) + x_{0,1}x_{1,0}x_{0,1}x_{1,2}\\ &= (\sqrt{2}(2+3) + \sqrt{2})C^4 = 6\sqrt{2}C^4\end{aligned}$$

and

$$(x^4)_{0,4} = x_{0,1}x_{1,2}x_{2,3}x_{3,4} = \sqrt{24}C^4.$$

The second-order contribution is

$$-E_0^{(2)} = \frac{\alpha^2}{x_0^8\hbar\omega}\left\{\frac{|x_{02}^4|^2}{2} + \frac{|x_{04}^4|^2}{4}\right\} = \frac{\alpha^2C^8}{x_0^8\hbar\omega}\left\{\frac{|6\sqrt{2}|^2}{2} + \frac{|\sqrt{24}|^2}{4}\right\}.$$

The final result is

$$E_0 \approx \frac{1}{2}\hbar\omega + \frac{3}{4}\alpha - \frac{21\alpha^2}{8\hbar\omega},$$

and the second order is small if $\alpha \ll \hbar\omega$.

Problem 20.4 For an anharmonic oscillator with Hamiltonian $\hat{H} = (\hat{n} + \frac{1}{2})\hbar\omega + \alpha(\frac{x}{x_0})^3$, calculate the correction to the energy of the state n in first and second orders. Perturbation theory, however, cannot be sound. Why?

Solution 20.4

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega - \frac{15\alpha^2}{\hbar\omega}\left(n^2 + n + \frac{11}{30}\right).$$

However, the perturbed potential has no minimum and true discrete levels cannot exist.

Problem 20.5 A particle in the one-dimensional well $0 < x < L$ is perturbed by the small potential $V(x) = V_0 \cos(\frac{\pi x}{L})\theta(\frac{L}{2} - x)$. Calculate the first-order correction to the levels $n = 1$ and $n = 2$.

Solution 20.5 Using $\psi_n(x) = \sqrt{\frac{2}{L}} \sin(\frac{n\pi x}{L})$, we obtain

$E'_n = \frac{2V_0}{L} \int_0^{\frac{L}{2}} dx \sin(\frac{n\pi x}{L})^2 \cos(\frac{\pi x}{L}) = \frac{2V_0}{\pi} \int_0^{\frac{\pi}{2}} dt \sin(nt)^2 \cos(t)$. For $n = 1$, $E'_1 = \frac{2V_0}{3\pi}$, and for $n = 2$, $E'_2 = \frac{8V_0}{\pi} \int_0^{\frac{\pi}{2}} d(t) \sin(t)^2 \cos^3(t)$. Integrating in $d\sin(t)$, one finds $= 8\frac{V_0}{\pi} \int_0^1 dx(x^2 - x^4) = \frac{16V_0}{15\pi}$.

20.1.2 Perturbation Theory for Degenerate Levels

Let the eigenvalue $E_n^{(0)}$ be g times degenerate; let

$$\{\psi_\nu^{(0)}\}, \quad \nu = 1, \ldots, g \tag{20.12}$$

denote an orthogonal basis for the subspace of degenerate eigenfunctions of H_0. At order 0, Eq. (20.4) reduces to $(E_n^{(0)} - E_k^{(0)})c_k = 0$, where k denotes an unperturbed level outside the subspace. Hence, $c_k^{(0)} = 0$ for $E_n^{(0)} \neq E_k^{(0)}$, that is , there is no mixing.

First Order

Therefore, to calculate the first-order corrections E_n^1, we ignore the mixing with the eigenfunctions outside the subspace. In the subspace, however, the Hamiltonian

$$H_{\mu,\nu} = E_n^{(0)}\delta_{\mu,\nu} + V_{\mu,\nu}, \quad \mu, \ \nu = 1 \ldots g, \tag{20.13}$$

is not special in any way and instead of an explicit expression we have $\hat{V}$ to diagonalize . Letting $E = E_n^0 + E_n^1$, we must solve the secular equation

$$boxed E_n^1 c_\mu = \sum_\nu V_{\mu,\nu} c_\nu. \tag{20.14}$$

The eigenvectors are the eigenfunctions of the 0-th approximation, which are linear combinations of the unperturbed $\psi_\nu^{(0)}$.

Second Order

The levels outside the subspace (20.12) come into play the second and higher approximations. Usually, this is important when the degeneracy is not resolved in the first approximation.

In such cases, to find the correction $E^{(2)}$, one starts from

$$(E - E_n^{(0)})c_\mu \equiv E^{(2)}c_\mu = \sum_m V_{\mu m}c_m^{(1)} \tag{20.15}$$

where $c_m^{(1)}$ is the amplitude of the state m outside the degenerate space, and is determined by

$$(E_n - E_m)c_m = \sum_\nu V_{m\nu}c_\nu, \tag{20.16}$$

which is accurate to the first order. Substituting,

$$E^{(2)}c_\mu = \sum_m \frac{V_{\mu m}}{E_\mu - E_m} \sum_\nu V_{m\nu}c_\nu.$$

Therefore, the second-order correction is given by condition:

$$det\left| \sum_m \frac{V_{m\mu}V_{\nu m}}{E_\mu - E_m} - E^2\delta_{\mu\nu} \right| = 0. \tag{20.17}$$

So it is still a secular equation. The change with respect to the previous case (20.14) is $V_{\mu\nu} \to \sum_m \frac{V_{m\mu}V_{\nu m}}{E_\mu - E_m}$. The subspace is not coupled to the other states directly but interacts in second order by a virtual excitation.

Problem 20.6 Recall (from Sect. 13.5) that for the plane rigid rotor, with momentum of inertia I and Hamiltonian $H = \frac{\hat{L}_z^2}{2I}$, the eigenfunctions are $\psi_k(\phi) = \frac{e^{ik\phi}}{\sqrt{2\pi}}$ with eigenvalues $k = 0$, and $k = \pm m\hbar, m = 1, 2, \ldots$. The energy eigenvalues $E_k = \hbar^2\frac{k^2}{2I}$ are degenerate, except $k = 0$. Find the effect of the directional perturbation

$$\hat{V} = 2\pi\lambda\,\delta(\phi - \phi_0),$$

where λ, ϕ_0 are constants.

Solution 20.6 The ground level $k = 0$ gets shifted at first order in λ. For the excited states,

$$\langle\phi_m|V|\phi_m\rangle = \lambda\int_0^{2\pi} d\phi e^{-im\phi}\delta(\phi - \phi_0)e^{im\phi} = \lambda$$

$$\langle\phi_{-m}|V|\phi_m\rangle = \lambda\int_0^{2\pi} d\phi e^{im\phi}\delta(\phi - \phi_0)e^{im\phi} = \lambda e^{2im\phi_0},$$

and the perturbation mixes the functions i $\psi_{-m}, \psi_m,\ m > 0$. The matrix reads as

$$V = \begin{bmatrix} \lambda & \lambda e^{2im\phi_0} \\ \lambda e^{-2im\phi_0} & \lambda \end{bmatrix}.$$

The eigenvalues give the first-order corrections.

The eigenvalue $E^1 = 0$ corresponds to the eigenvector $\frac{1}{\sqrt{2}}\begin{pmatrix} -e^{im\phi_0} \\ e^{-im\phi_0} \end{pmatrix}$ of V, which is

$$\psi_- = -\frac{e^{im\phi_0}}{\sqrt{2}}\phi_{-m} + \frac{e^{-i\ \mathrm{m}\phi_0}}{\sqrt{2}}\phi_m = i\frac{\sin\left[m(\phi-\phi_0)\right]}{\sqrt{\pi}}$$

and is not affected by the perturbation. The eigenvalue $E^1 = 2\lambda$ corresponds to the eigenvector $\frac{1}{\sqrt{2}}\begin{pmatrix} e^{im\phi_0} \\ e^{-im\phi_0} \end{pmatrix}$, that is, to $\psi_+ = \frac{\cos[m(\phi-\phi_0)]}{\sqrt{\pi}}$.

Problem 20.7 Explain the *Stark–Lo Surdo effect* for the H atom with $n = 2$ (the H atom plunged in an electric field $\mathcal{E}$).

Solution 20.7 The perturbation

$$V = ez\mathcal{E}$$

acts in first order *(linear Stark–Lo Surdo effect)*. The wave functions $\psi_{nlm} = R_{nl}Y_{lm}(\theta,\phi)$ involved are

$$\begin{cases} \psi_{200} & R_{20}(r) = 2(\frac{1}{2a_0})^{\frac{3}{2}}(1-\frac{r}{2a_0})e^{-\frac{r}{2a_0}} & Y_{00} = \frac{1}{\sqrt{4\pi}}, \\ \psi_{210} & R_{21}(r) = \frac{1}{\sqrt{3}}(\frac{1}{2a_0})^{\frac{3}{2}}\frac{r}{a_0}e^{-\frac{r}{2a_0}} & Y_{10} = \sqrt{\frac{3}{4\pi}}\cos(\theta), \\ \psi_{211} & R_{21}(r) = \frac{1}{\sqrt{3}}(\frac{1}{2a_0})^{\frac{3}{2}}\frac{r}{a_0}e^{-\frac{r}{2a_0}} & Y_{11} = -\sqrt{\frac{3}{8\pi}}\sin(\theta)e^{i\phi}, \\ \psi_{21-1} & R_{21}(r) = \frac{1}{\sqrt{3}}(\frac{1}{2a_0})^{\frac{3}{2}}\frac{r}{a_0}e^{-\frac{r}{2a_0}} & Y_{1-1} = -\sqrt{\frac{3}{8\pi}}\sin(\theta)e^{-i\phi}. \end{cases}$$

We need the matrix of the perturbation on this basis. Since V is odd, the diagonal elements vanish, and since V does not depend on ϕ, the matrix elements between functions with different m also vanish. ψ_{200} gets mixed with ψ_{210}, and effectively the matrix that we must diagonalize is 2×2. Using $\cos(\theta) = \frac{z}{r}$, one finds

$$\langle\psi_{200}|V|\psi_{210}\rangle = 3ea_0\mathcal{E}$$

and

$$\begin{cases} E^{(1)}_- = 3ea_0\mathcal{E} & \frac{\psi_{200}-\psi_{210}}{\sqrt{2}} \\ E^{(1)}_+ = -3ea_0\mathcal{E} & \frac{\psi_{200}+\psi_{210}}{\sqrt{2}} \end{cases}.$$

The degenerate level splits into 3.

20.2 Time-Dependent Perturbations

Transitions between quantum states occur under the influence of some time-dependent external perturbation $\hat{V}(t)$. This is what one normally does to study a system. In order to test the properties of a molecule or a solid, one uses a probe like a weak field that produces excitations, but still allows us to classify the states in terms of the stationary states of the unperturbed system, while allowing transitions between them. The present section is therefore fundamental for all spectroscopies. We wish to solve

$$i\hbar\frac{\partial\psi}{\partial t} = \left[H_0 + \hat{V}(t)\right]\psi \tag{20.18}$$

in terms of the full orthonormal set of solutions of the *unperturbed* problem

$$i\hbar\frac{\partial\phi_n}{\partial t} = H_0\phi_n(t), \tag{20.19}$$

evolving according to

$$\phi_n(t) = \phi_n(0)\exp\left(\frac{-i\epsilon_n t}{\hbar}\right). \tag{20.20}$$

Let us suppose that the system described by H_0 is prepared in the eigenstate ϕ_n, but under the action of the *weak* perturbation $\hat{V}(t)$, it makes a transition to other unperturbed states. First, let us expand[1] on the unperturbed basis with $\psi = \sum_k a_k(t)\phi_k(t)$; substituting into (20.18), we find

$$i\hbar\sum_k \dot{a}_k(t)\phi_k(t) = \sum_k a_k(t)\hat{V}\phi_k(t).$$

Scalar multiplication by $\phi_m(t)$ gives us

$$i\hbar\dot{a}_m = \sum_k V_{mk}(t)a_k(t), \tag{20.21}$$

where

$$V_{mk}(t) = \langle\phi_m(t)|\hat{V}(t)|\phi_k(t)\rangle = \langle\phi_m(0)|\hat{V}(t)|\phi_k(0)\rangle e^{i\omega_{mk}t}, \tag{20.22}$$

$$\hbar\omega_{mk} = \epsilon_m - \epsilon_k. \tag{20.23}$$

Note the double time dependence of $V_{mk}(t)$ through the operator and the phase factor. From the *initial conditions*

$$\psi(0) = \phi_n(0) \implies a_k(0) = \delta_{kn}, \tag{20.24}$$

[1] P. A. M. Dirac, 1926.

we can write an exact formal solution, in the presence of $\hat{V}$ and obtain the transition amplitude to any final state m. For clarity,

we shall denote the amplitude to go to final state m by $a_{n\to m}$.

In view of (20.23), if $\omega_{kn} > 0$, the system is *promoted* to a higher level m, while for $\omega_{mn} < 0$, we speak of a *decay* to a lower level. The system energy is not conserved, since H is time-dependent.

So far we have made no approximations, but usually the exact solution is prohibitively difficult or expensive. Now we introduce the assumption that the perturbation is so weak that the probability of exciting the system is small. In this case, we feel licensed not only to use perturbation theory, but even to stop after one step, or *first approximation*. Simply, we put into the R.H.S. of (20.21) the initial condition (20.24) $\psi(t) \sim \phi_n(t)$, or, in other words, $a_k(0) \sim \delta_{kn}$. So,

$$i\hbar \frac{d}{dt} a_{n\to m}(t) = V_{mn}(t),$$

which is solved by

$$a_{n\to m}(t) = -\frac{i}{\hbar} \int_{t_{min}}^{t_{max}} V_{mn}(\tau) d\tau. \tag{20.25}$$

The integration limits t_{min}, t_{max} depend on the problem. For instance, if $\hat{V}$ is switched on at time 0 and off at a later time T, we write $\int_0^T$, and so on.

Problem 20.8 A particle is confined between $x = 0$ and $x = a$ in a deep one-dimensional well. At time $t = 0$, it is in the ground state.

a. Calculate the first-order transition probability P_m to the m-th excited state under the action of a potential $V(t) = \lambda\delta(x - \frac{a}{2})\theta(t)\theta(T - t)$.
b. State the validity criterion of the approximation.
c. Find T such that the transition probability to the second excited level $m = 3$ is maximum.

Solution 20.8 Using the eigenfunctions (11.3) $\psi_n(x) = \sqrt{\frac{2}{a}} \sin\left[n\pi \frac{x}{a}\right]$, one finds

$$V_{m1}(t) = \frac{2\lambda}{a} \int dx \sin\left(\frac{m\pi x}{a}\right) \sin\left(\frac{\pi x}{a}\right) \delta\left(x - \frac{a}{2}\right) e^{i\omega_{m1}\tau},$$

with

$$\omega_{m1} = \frac{\hbar\pi^2(m^2 - 1)}{2ma^2}.$$

Thus, $P_m = |\frac{2\lambda}{\hbar a} \sin(m\pi/2) \int_0^T e^{i\omega_{m1}\tau} d\tau|^2$. Moreover, $P_m \ll 1$.
Since $|\int_0^T e^{i\omega_{m1}\tau} d\tau|^2 = 2\frac{1-\cos(\omega_{m1}T)}{\omega_{m1}^2}$, the condition is $\cos(\omega_{31}T) = -1$, which implies that $\omega_{31}T = \pi, 3\pi, 5\pi, \ldots$

Problem 20.9 A harmonic oscillator is in the ground state at time $t = 0$. Calculate the probability of transition to level n at time T under the action of the weak perturbation $V(t) = (a + a^\dagger)V_0 T\delta(T - t)$.

Solution 20.9 It is only possible to jump to $n = 1$. The probability is $|\frac{V_0 T}{\hbar}|^2$.

20.2.1 Periodic Perturbations and Fermi Golden Rule

The interaction of charges with a monochromatic electromagnetic wave is most often well described in lowest order perturbation theory, unless one wants to study nonlinear optics by means of intense laser radiation. In most cases, one introduces the minimal coupling rule (2.94), neglecting the $\overrightarrow{A}^2$ term on the grounds that it is quadratic. Then setting $\boldsymbol{A}(t) = \boldsymbol{A}_0 e^{i\omega t} + \boldsymbol{A}_0^* e^{-i\omega t}$, one considers a periodic perturbation

$$\hat{V}(t) = \hat{W} e^{i\omega t} + \hat{W}^\dagger e^{-i\omega t}, \quad \omega > 0, \tag{20.26}$$

where $\hat{W} = \overrightarrow{A}_0 \cdot \overrightarrow{p}$. We shall see that the first term produces emission and the second absorption of a photon. Since the integrand is periodic, we use (20.25) with integration limits $-\infty, \infty$. However, working out

$$a_{n\to m}(t) = -\frac{i W_{mn}}{\hbar} \int_{-\infty}^{\infty} \left[e^{i(\omega_{mn}+\omega)\tau} + e^{i(\omega_{mn}-\omega)\tau} \right] d\tau,$$

and using $\int_{-\infty}^{\infty} e^{imx} = 2\pi\delta(x)$, we get this result:

$$a_{n\to m} \equiv a_m(t \to \infty) \stackrel{?}{=} -\frac{i W_{mn}}{\hbar} 2\pi \left[\delta(\omega_{mn} + \omega) + \delta(\omega_{mn} - \omega) \right].$$

From one point of view this is reasonable. If $\omega_{mn} < 0$ (emission), only the first term contributes, if $\omega_{mn} > 0$ (absorption) only the second. In this way, one obtains delta-like lines with the resonance condition $\pm\omega_{mn} = \omega$ for absorption and emission.

However, the result is wrong. This is a transition amplitude, and in order to find the probability $P_{n\to m} = a^*_{n\to m} a_{n\to m}$, we must square; alas, $\delta(x)^2$ is a mathematical nonsense. We made a mistake, but where?

Actually, it is not a mathematical error, but an oversimplified scheme. A perfectly monochromatic excitation should last for a time $T \to \infty$ and cannot exist. Often in Physics problems, one avoids trouble by introducing a cutoff, that is, by replacing zero with a small quantity and infinity with a large one. Even here, we get out of trouble if we introduce the duration T of the excitation. For the absorption of a wave that lasts a time T, we find

$$a_{n\to m}(t) = -\frac{i W_{mn}}{\hbar} \int_0^T e^{i(\omega_{mn}-\omega)\tau} d\tau,$$

and so

$$P_{n\to m} = \frac{iW^*_{mn}}{\hbar}\int_0^T e^{-i(\omega_{mn}-\omega)\tau_1}\, d\tau_1 \times \left(-\frac{iW_{mn}}{\hbar}\right)\int_0^T e^{i(\omega_{mn}-\omega)\tau_2} d\tau_2.$$

Since $\int_0^T e^{ix\tau}d\tau = \frac{e^{ixT}-1}{ix}$, now the lines have some width $\sim \frac{1}{T}$. The problem of the squared delta is solved. Then, we can decide to neglect the finite width and simplify the result by approximating one of the integrals with $\int_0^T \sim \int_{-\infty}^{\infty} = 2\pi\delta\left(\omega_{mn}-\omega\right)$; then, the other integral equals T. In this way,

$$P_{n\to m} = \left|\frac{W_{mn}}{\hbar}\right|^2 T \times\ 2\pi\delta\left(\omega_{mn}-\omega\right).$$

This probability grows linearly with the duration T of the experiment. With increasing T, one first violates the condition $P \ll 1$ which is needed for first-order perturbation theory, and eventually one finds $P > 1$ which makes no sense. One solution would be to evaluate a realistic duration T and plug it into the formula. Actually, one can avoid this step, by defining the transition rate

$$R_{n\to m} = \frac{\partial P_{n\to m}}{\partial T}$$

which is the actually measured quantity. In this way, one arrives at the *Fermi golden rule*:

$$R_{n\to m} = \frac{1}{\hbar}\left|W_{mn}\right|^2 2\pi\delta\left(\epsilon_m-\epsilon_n-\hbar\omega\right). \tag{20.27}$$

Problem 20.10 Let $H_0 = \epsilon\sigma_z$; $\hat{V} = \lambda(e^{i\omega t}+e^{-i\omega t})\sigma_x$, with $\psi(0) = \binom{0}{1}$; calculate the rate R.

Solution 20.10 $R = \frac{2\pi\lambda^2}{\hbar}\delta(2\epsilon-\hbar\omega)$.

Selection Rules for the H Atom

The H atom can absorb a photon and promote the electron to a higher level; when excited, it emits a photon while the electron jumps down to a lower level. The transition rate can be obtained in the dipole approximation assuming a uniform electric field; this is good in the visible or near ultraviolet, since the wavelength is on the order of hundreds of nanometers and is large compared to the Bohr[2] radius. To find

[2] Niels Henrik David Bohr was a Danish physicist who made foundational contributions to understanding atomic structure and quantum theory. Bohr's 1913 model of the atom is still popular although it is really rather misleading, but he received the Nobel Prize in Physics in 1922 for his subsequent work on atomic structure. Bohr was also a philosopher and a promoter of scientific research.

the transition rate, one computes the matrix elements of $\overrightarrow{r} = (x, y, z)$ between the initial and final wave functions. The transition occurs if at least one between

$$A_x = \langle \psi_{n'l'm'} | x | \psi_{nlm} \rangle,\ A_y = \langle \psi_{n'l'm'} | y | \psi_{nlm} \rangle,\ A_z = \langle \psi_{n'l'm'} | z | \psi_{nlm} \rangle$$

does not vanish. Since the dipole is odd, this requires that the initial and final wave functions have opposite parity. Since the parity of the harmonic $Y_{lm}(\theta, \phi)$ is $(-1)^l$, this requires that l e l' have opposite parity (*Laporte's rule*).

Besides the parity, the total angular momentum (including the photon) is conserved. This leads to the rules

$$\boxed{m' = m \text{ or } m' = m \pm 1, l' = l \pm 1.} \tag{20.28}$$

These rules forbid the transitions for which the matrix element vanishes. Indeed, up to a constant, $z \sim Y_{10}(\theta, \phi)$, $(\mathrm{x} \pm iy) \sim Y_{1\pm1}(\theta, \phi)$. Thus the angular factor in the matrix element reads as

$$\int d\Omega Y^*_{l'm'}(\theta, \phi) Y_{1q}(\theta, \phi) Y_{lm}(\theta, \phi) = \sqrt{\frac{3}{4\pi} \frac{2l+1}{2l'+1}} \langle l100 | l'0 \rangle \langle l1mq | l'm' \rangle,$$

and the forbidden transitions depend on vanishing Clebsh–Gordan coefficients. The result shows that the Photon spin is 1.

20.3 The Evolution Operator

It is time to consider more advanced techniques that allow a general approach to perturbation theory. For the sake of generality, let us consider a time-dependent system with Hamiltonian

$$\hat{H}(t) = \hat{H}_0 + \hat{V}(t), \tag{20.29}$$

where $\hat{H}_0$ is static and easy to deal with, while $\hat{V}(t)$ is a perturbation which is not assumed to be small and is not necessarily time-dependent. Moreover, $\hat{H}_0$ and $\hat{H}(t)$ are in the Schrödinger picture (the one that we have used up to now); to stress this fact, we may write

$$i \frac{\partial}{\partial t} |\Psi_S(t)\rangle = \hat{H}_S(t) |\Psi_S(t)\rangle. \tag{20.30}$$

A formal solution is

$$|\Psi(t)\rangle = U_S(t, t_0) |\Psi(t_0)\rangle, \tag{20.31}$$

where $U_S(t, t_0)$ is the evolution operator from the "initial" time t_0. This is clever, not only for the general case of time-dependent problems, but even when the physical problem is time-independent, because we can pick t_0 in the far past such that

$\hat{V}(t < t_0) = 0$ and $\Psi(t < t_0) = 0$ corresponds to an eigenstate of $\hat{H}_0$ with eigenvalue E_0. Even in stationary problems, when V is constant, we shall pretend that V was adiabatically switched on starting with a golden age $t_0 \to -\infty$ when the Hamiltonian was H_0 and life was easy. From (20.31)), one finds that the evolution operator is unitary and solves

$$i\frac{\partial}{\partial t}U_S(t,t_0) = \hat{H}_S U_S(t,t_0). \tag{20.32}$$

Integrating,

$$U_S(t,t_0) = 1 - i\int_{t_0}^{t} dt_1 H_S(t_1)U_S(t_1,t_0); \tag{20.33}$$

solving by iteration,

$$U_S(t,t_0) = 1 - i\int_{t_0}^{t} dt_1 H_S(t_1) + (-i)^2\int_{t_0}^{t} dt_1\int_{t_0}^{t_1} dt_2 H_S(t_1)H_S(t_2) + \cdots \tag{20.34}$$

The n-th term is

$$(-i)^n\int_{t_0}^{t} dt_1\int_{t_0}^{t_1} dt_2\cdots\int_{t_0}^{t_{n-1}]-} dt_n H_S(t_1)H_S(t_2)\cdots H_S(t_n)$$

$$= (-i)^n\int_{t_0}^{t} dt_1\int_{t_0}^{t_1} dt_2\cdots\int_{t_0}^{t_{n-1}]-} dt_n H_S(t_1)H_S(t_2)\cdots H_S(t_n)\theta(t_1-t_2)\cdots\theta(t_{n-1}-t_n).$$

Any of the n! permutations of the labels 1,2,$\cdots$ n transforms this integral into an identical integral with the times in decreasing order. Thus we can sum over the permutations and divide by n!. Let P denote the time ordering operator defined by

$$P[H_S(t_1)H_S(t_2)\cdots H_S(t_n)] = H_S(t_1)H_S(t_2)\cdots H_S(t_n)\theta(t_1-t_2)\cdots\theta(t_{n-1}-t_n)$$

$$+ H_S(t_2)H_S(t_1)\cdots H_S(t_n)\theta(t_2-t_1)\cdots\theta(t_{n-1}-t_n) + \cdots$$

Most often one uses the Wick cronological operator, which is like P except that any exchange of two Fermion creation or annihilation operators produces a minus sign (that is, the Fermion operators anticommute). This simplifies the definition of Fermion Green's functions in many-body problems. Now we can sum (at least formally) the series:

$$U_S(t,t_0) = Texp(-i\int_{t_0}^{t} dt' H_S(t')). \tag{20.35}$$

20.4 Heisenberg Representation and Finite Temperature

The many-body Green's functions are defined as expectation values in the Heisenberg[3] representation, in which the wave function $|\Psi_H\rangle \equiv |\Psi_S(t_0)\rangle$ does not depend on time but is taken to be the usual Schrödinger wave function photographed at some fixed time t_0. The expectation values are written as

$$\langle A(t)\rangle = \langle \Psi_H | A_H(t) | \Psi_H \rangle. \tag{20.36}$$

All the time dependence is taken care of by the operators:

$$A_H(t) = U_S^\dagger(t, t_0) A_S U_S(t, t_0). \tag{20.37}$$

Let $\Psi_\alpha^\dagger$ be the creation operator in a state defined by a set α of quantum numbers. The one particle properties can be studied in terms of the one-particle Green's functions

$$g^<(t, t') = \langle \Psi_H^\dagger(t') \Psi_H(t) \rangle \tag{20.38}$$

$$g^>(t, t') = \langle \Psi_H(t) \Psi_H^\dagger(t') \rangle. \tag{20.39}$$

In equilibrium, $g^<$ accounts for the occupied states and $g^>$(for the empty ones. It is a simple exercise to show that if $\hat{A}$ is a time-independent operator,

$$i \frac{d\hat{A}_H}{dt} = [\hat{A}_H, H_H]_-. \tag{20.40}$$

The Heisenberg picture is often used to exploit the formal analogy of the complex time dependence of energy eigenstates going like $e^{\frac{-iEt}{\hbar}}$ with the Boltzmann factor. The thermal average of an operator $\hat{A(t)}$ at temperature T is given by

$$\langle A(t)\rangle = Tr \rho_0 A_H(t), \tag{20.41}$$

[3] Werner Karl Heisenberg was a German theoretical physicist and one of the main pioneers of the theory of quantum mechanics. He published his work in 1925 in a major breakthrough paper. Heisenberg was awarded the Nobel Prize in Physics in 1932 "for the creation of quantum mechanics". He is also known for his work on the uncertainty principle which he introduced in 1927. Heisenberg made important contributions to the theories of turbulent flows, atomic nuclei, ferromagnetism, cosmic rays and subatomic particles. During World War II, he was one of the main scientists behind Germany's nuclear weapons program. After the war, he became a vocal advocate for nuclear disarmament and international cooperation in science.

with

$$\rho_0 = \frac{e^{-\beta H_0}}{Tr(e^{-\beta H_0})} \tag{20.42}$$

with $\beta = \frac{1}{K_B T}$.

20.5 Interaction Representation

The Heisenberg representation is important conceptually and the operators that appear in Green's function approach to many-body interacting problems are Heisenberg operators. However, the main strategy for obtaining numbers is the perturbation approach, which is an expansion in powers of the interaction $\hat{V}(t)$, and the Heisenberg representation does not provide a viable strategy. To this end, P. A. M. Dirac introduced the interaction picture (or representation) in which any (possibly time-dependent) operator $\hat{A}_S(t)$ becomes A_I and evolves as follows:

$$A_I = e^{iH_0 t} A_S e^{-iH_0 t}, \tag{20.43}$$

while the Schrödinger wave function $|\Psi_S(t)\rangle$ is related to the interaction wave function, which is

$$|\Psi_I(t)\rangle = e^{iH_0 t}|\Psi_S(t)\rangle, \tag{20.44}$$

and

$$\langle \hat{A}_S(t)\rangle = \langle \Psi_S(t)|\hat{A}_S(t)|\Psi_S(t)\rangle = \langle \Psi_I(t)|\hat{A}_I(t)|\Psi_I(t)\rangle. \tag{20.45}$$

Now,

$$i\frac{\partial}{\partial t}|\Psi_I(t) = V_I(t)|\Psi_I(t)\rangle \tag{20.46}$$

and we face the same problem as in equations (20.31, 20.32) but with $V_I(t)$ instead of $U_S(t, t_0)$, so we can repeat the argument. By the way, it is a good idea to start the integration again at a time t_0 and to introduce the evolution operator in the interaction picture

$$U_I(t, t_0) = Texp\left[-i\int_{t_0}^{t} dt' V_I(t')\right]. \tag{20.47}$$

This is the starting point of all the perturbation expansions. Indeed, a successful strategy is the following. One assumes that t_0 belongs to the remote past, and that for $t < t_0$, there was a *golden age* during which $V = 0$ and life was easy for theoreticians. So one can evolve the system starting with the ground state of H_0, which is really a wave packet involving all the excited states of H. However, one can justify the assumption that during the long time t_0 the excited states decay (when doing the zero

temperature theory) and/or the system goes to an equilibrium (in finite temperature theory). So, this is the starting point of many-body theory.

The path from the golden age to time t can be generalized in order to deal with time-dependent Hamiltonians and finite temperatures in the Keldysh formalism, which is one of the advanced tools of many-body theory and numerical simulations.

20.5.1 From Heisenberg Picture to Interaction Picture

In the Heisenberg representation (20.37), $\langle A(t)\rangle = \langle \Psi_H | A_H | \Psi_H \rangle$, where $|\Psi_H\rangle \equiv |\Psi_S(t_0)\rangle$ and t_0 is an arbitrary reference time. In the interaction picture,

$$\Psi_I(t_0\rangle = e^{iE_0 t_0}\Psi_S(t_0),$$

therefore

$$\langle A(t)\rangle = \langle \Psi_I(t_0) A_H(t) | \Psi_I(t_0)\rangle \tag{20.48}$$

but also

$$\langle A(t)\rangle = \langle \Psi_I(t_0) A_I(t) | \Psi_I(t_0)\rangle. \tag{20.49}$$

Therefore,

$$A_H(t) = U_I^\dagger(t, t_0) A_I U_I(t, t_0). \tag{20.50}$$

Thus the Heisenberg operators can be expanded in powers of the interaction.

Chapter 21
Variational Principle for Schrödinger–Pauli Theory

Variational principles are ubiquitous in Theoretical Physics and are very useful mathematical tools. This one depends on the existence of a ground state.

21.1 The Ground State and the Absolute Minimum of Energy

The Schrödinger–Pauli theory is characterized by the fact that every system must have a ground state, whose energy is a lower bound to the energies of all states. This is not true in Classical Mechanics, in which, for instance, a H atom could have any energy.

Let ϕ denote any wave function of an arbitrary quantum system, with one or several particles, duly normalized with

$$N =< \phi|\phi >= 1. \tag{21.1}$$

The expectation value of the energy

$$E =< \phi|H|\phi > \tag{21.2}$$

cannot be lower than the energy ϵ_0 of the ground state. To see that, let us expand ϕ in eigenvectors ψ_n of H:

$$|\phi\rangle = \sum_n |\psi_n\rangle\langle\psi_n|\phi\rangle.$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_21

Since $\epsilon_n \geq \epsilon_0$,

$$E = \sum_n |\langle \psi_n | \phi \rangle|^2 \epsilon_n \geq \epsilon_0. \tag{21.3}$$

So, the ground state can be sought by looking for the energy minimum. We can get close to the real ground state if we start from a correct qualitative idea of the form of the ground wave function. To explore this possibility, let us consider a familiar case, namely the ground state of the harmonic oscillator ($H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2$). We can pick the exact solution among the normalized trial functions, depending on the unknown parameter b,

$$\phi(x) = Ae^{-bx^2}, \quad A = \left(\frac{2b}{\pi}\right)^{1/4}.$$

We use the familiar integrals $\int_\infty^\infty dx \exp\left[-\alpha x^2\right] = \sqrt{\frac{\pi}{\alpha}}$, $\int_\infty^\infty dx x^2 \exp\left[-\alpha x^2\right] = \frac{1}{2}\sqrt{\frac{\pi}{\alpha^3}}$. Using $\langle \phi | x^2 | \phi \rangle = \frac{1}{4b}$, $\langle \phi | \frac{d^2}{dx^2} | \phi \rangle = \langle \phi | (-2b + 4b^2 x^2) | \phi \rangle = -b$, one finds that

$$E = \frac{\hbar^2 b}{2m} + \frac{m\omega^2}{8b}.$$

The first term is kinetic energy and the second is potential energy. $E(b)$ has one minimum. $\frac{dE}{db} = \frac{\hbar^2}{2m} - \frac{m\omega^2}{8b^2} = 0$ yields $b = \frac{m\omega}{2\hbar}$, $E = \hbar\omega/2$. In this way, one arrives at the exact solution, since the solution was one of the trial functions.

The energy (21.2) is a quadratic *functional*[1] of ϕ. Let us see the effect on E of a variation of ϕ; the variation must be infinitesimal but otherwise arbitrary. We shall denote the variation as $\phi \to \phi + \alpha\eta$, where η is an arbitrary complex function of the same variables as ϕ, while $\alpha \to 0$ is a complex parameter with arbitrary phase. The variation of energy $E =< \phi | H | \phi >$, which follows from $\phi \to \phi + \alpha\eta$, in first order in α, is therefore

$$\delta E = \alpha^* < \eta | H | \phi > + \alpha < \phi | H | \eta > + O(\alpha^2),$$

where $O(\alpha^2)$ is negligible. The condition for a stationary E is

$$E \quad \text{is stationay} \iff \{\delta E = 0, \quad \text{arbitrary} \quad \eta\}. \tag{21.4}$$

The condition (21.4) is too strong to be interesting, since we cannot accept the variations of ϕ that violate the normalization condition. The variation changes the norm by

$$\delta N = \alpha^* < \eta | \phi > + \alpha < \phi | \eta > + O(\alpha^2).$$

[1] A functional of ϕ is an integral that has ϕ in the integrand; it can be considered as a function of infinitely many variables, which are the values taken by ϕ in the field of integration. This one is quadratic, since ϕ appears in bra and ket.

Therefore, the extremum must be conditioned. Using the Lagrange method we write

$$\delta(E - \lambda N) = 0.$$

Note that this condition is obtained from the unconstrained minimum condition $\delta E = \delta\langle\phi|H|\phi\rangle = 0$ by the substitution $H \to H - \lambda$. In this way, $|\phi >$ depends on the multiplier λ, which is fixed eventually by setting $N =< \phi(\lambda)|\phi(\lambda) >= 1$.

We obtain the same result more simply if we vary only the bra, that is redefine δE and δN as

$$\delta E = \alpha^* < \eta|H|\phi >, \quad \delta N = \alpha^* < \eta|\phi > . \tag{21.5}$$

By the Lagrange method, we obtain the constrained minimum condition

$$\alpha^* < \eta|H - \lambda|\phi >= 0.$$

Since η is arbitrary, it follows that

$$(H - \lambda)|\phi >= 0.$$

Thus, the condition coincides with the stationary state equation and λ coincides with the eigenvalue E. In summary,

$$\{H\phi = E\phi, \ < \phi|\phi >= 1\} \Leftrightarrow \{\delta(E - \lambda N) = 0, \lambda = E\} \ \Leftrightarrow \{\delta(E) = 0, \ N = 1\}.$$

The extremum condition is another way to state the stationary state Schrödinger equation. It holds for all the eigenstates, and the Lagrange multiplier is the energy eigenvalue.

Example

Consider the Hamiltonian matrix

$$H = \begin{pmatrix} -3 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Use the variational method to find the eigenvectors of the following form:

$$\psi = \begin{pmatrix} \alpha \\ \beta \\ \beta \\ \beta \\ \beta \end{pmatrix}.$$

The normalization requires $N = \langle\psi|\psi\rangle = \alpha^2 + 4\beta^2 = 1$; the energy is $E = \langle\psi|H|\psi\rangle = 8\alpha\beta - 3\alpha^2$: the principle requires that we extremize $f(\alpha, \beta) = E - \lambda N$. One finds

$$\begin{cases} \frac{\partial f}{\partial \alpha} = 0 \Longrightarrow 4\beta = (3+\lambda)\alpha \\ \frac{\partial f}{\partial \beta} = 0 \Longrightarrow \alpha = \lambda\beta. \end{cases}$$

The condition $\lambda(3+\lambda) = 4$ gives the roots $\lambda = -4$, $\lambda = 1$. For $\lambda = -4$ from $\alpha = -4\beta$, we get

$$\psi_{-4} = \frac{1}{\sqrt{20}} \begin{pmatrix} -4 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

which is the exact ground state with eigenvalue $\epsilon = -4$.

For $\lambda = 1$, $\alpha = \beta$ gives us $\psi_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$, which is the exact excited state with eigenvalue $\epsilon = 1$.

21.2 Variational Approximations for One-Particle Problems

The great theoretician Richard Feynman wrote that, in order to make progress, a theoretical physicist must understand a given involved problem not is one but in several different ways. The variational principle is equivalent to the stationary state Schrödinger equation, and aside from the mathematical interest, it is a very useful tool, for generating approximations of otherwise intractable problems. The real power of this method has been demonstrated by its application to realistic hard many-body problems, but in the present subsection we concentrate on one-particle examples.

One can invent trial functions $\phi(x, \{\lambda_1, \lambda_2, \ldots \lambda_n\})$ that depend on a number of parameters $\{\lambda_1, \lambda_2, \ldots \lambda_n\}$. Then, the method requires a minimization of the energy as a function of the parameters. If ϕ is not already normalized, the normalization constraint can be imposed via a Lagrange multiplier. If the exact ground ϕ happens to be in the class of trial functions, the exact result is gained. Otherwise, it is good to know that the approximate E is higher than the exact value. The approximation gets better if we improve or enlarge the class of functions and increase the number of parameters. Typically, one obtains accurate estimates of E, even with relatively poor $\phi(x)$.

Even the excited states correspond to extrema of the functional, but the method has a limited applicability to them.

> *The trouble is that the true eigenstates must be orthogonal, but this property cannot be granted for the f that belong to a limited class of functions.*

The orthogonality is granted in the LCAO (= linear combination of atomic orbitals) method for finding the molecular orbitals in a simple independent-particle scheme, but this is an exception. We cannot dispense from orthogonality; if an excited state is not orthogonal to the ground state, we cannot give it any meaning and if we orthogonalize, the variational principle is not satisfied. In general this method is not suited to excited states. However, the lowest state of each symmetry type can be sought through the variational method. For instance, the non-relativistic atomic states are labeled by L, m (angular momentum and its z component); states with different angular momenta are automatically orthogonal. So, if, for instance, the approximate ground state has $L = 0$, and the approximate first excited state has $L = 1$, we can use the variational principle in this case. Any symmetry works, since by symmetry, we mean operator X, which is unitary (that is $XX^\dagger = 1$; see Sect. 13.2), such that $[H, X]_- = 0$. Eigenstates of a unitary operator X belonging to different eigenvalues are orthogonal. Indeed, if $X\phi_1 = e^{i\alpha}\phi_1$ and $X\phi_2 = e^{i\beta}\phi_2$,

$$(\phi_1, \phi_2) = (\phi_1, X^\dagger X\phi_2) = e^{i(\beta-\alpha)}(\phi_1, \phi_2),$$

and, if $\alpha \neq \beta$, this requires that $(\phi_1, \phi_2) = 0$.

Problem 21.1 For the quartic anharmonic oscillator of Problem 36, find the variational condition for a trial function $\phi(x) = Ae^{-bx^2}$, $A = (\frac{2b}{\pi})^{1/4}$, using the identity $\int_{-\infty}^{\infty} dx x^4 \exp(-\alpha x^2) = \frac{3}{4}\sqrt{\frac{\pi}{\alpha^5}}$.

Solution 21.1 $E = \frac{\hbar^2 b}{2m} + \frac{m\omega^2}{8b} + \frac{3\alpha}{16x_0^4 b^2} = \text{minimum}$.

There is no general method for choosing the trial functions; a proper choice is important, and that is the really hard initial step of the variational theory. Some solutions are very famous, like the Hartree–Fock method; the theory of superconductivity by Bardeen, Cooper and Schrieffer was initially formulated in terms of a variational solution.

21.3 Hartree–Fock Method

The non-relativistic N-electron Hamiltonian for an atom, molecule or solid $H(r_1, r_2, \ldots, r_N) = H_0 + V$ consists of a free part and an interaction term. The free part is of the form

$$H_0 = T + H_w = \sum_i h(i) \equiv \sum_i \left\{ -\frac{1}{2}\nabla_i^2 + w(r_i) \right\}, \tag{21.6}$$

where $w(r_i)$ is the electrostatic potential of the nuclei acting on electron i. This part, taken alone, would be relatively easy to solve with N electrons, in terms of a determinantal state

$$\Psi(1, 2, \ldots N) = \frac{1}{\sqrt{N!}} \begin{pmatrix} u_1(1) & u_2(1) & \ldots & u_N(1) \\ u_1(2) & u_2(2) & \ldots & u_N(2) \\ \ldots & \ldots & \ldots & \ldots \\ u_1(N) & u_2(N) & \ldots & u_N(N) \end{pmatrix} \tag{21.7}$$

written in terms of spin-orbitals $u_i(i)$ to be determined. However, the result would be totally wrong physically. The great complication arises from the electron–electron interaction term

$$V = \frac{1}{2} \sum_{i \neq j} \frac{e^2}{|r_i - r_j|}, \tag{21.8}$$

which forces us to deal with all the electrons on equal footing. There is a dramatic increase in the dimensionality of the problem. The Hartree–Fock (HF) method, or mean field approximation, was conceived a century ago, when calculations were still done by a quill pen. It seeks to approximate the problem as a succession of partial subproblems in which one can deal with one electron at a time. In short, it seeks the best variational determinantal wave function. Using the rules of Chap. 25, the expectation value of H is found to be

$$E_N = \sum_i^N I_i + \frac{1}{2} \sum_{i \neq j}^N \left[C_{ij} - E_{ij} \right], \tag{21.9}$$

where $I_i = \langle u_i | h(i) | u_i \rangle$ and C_{ij}, E_{ij} denote the Coulomb and Exchange integrals. The first term

$$C_{ij} = \int d^3 r_1 d^3 r_2 \frac{|u_i(r_1)|^2 |u_j(r_2)|^2}{r_{12}} \tag{21.10}$$

has a clear electrostatic meaning, and is called the *Coulomb term*; the second contribution vanishes for anti-parallel spins, and for parallel spins it reads as the *exchange term*

$$E_{ij} = \int d^3 r_1 d^3 r_2 \frac{u_i^*(r_1) u_j^*(r_2) u_i(r_2) u_j(r_1)}{r_{12}}. \tag{21.11}$$

Looking for the extremum of energy constrained by normalization, one finds the Hartree–Fock (HF) equations. We introduce the direct potential

$$V^d = \sum_i^N V_i^d(\boldsymbol{r}),\ \ V_i^d(\boldsymbol{r}) = \int d\boldsymbol{r}'\, \frac{|u_i(\boldsymbol{r}')|^2}{|\boldsymbol{r} - \boldsymbol{r}'|}$$

summed over all electrons, and the exchange potential

$$V^{ex} = \sum_i^{\uparrow\uparrow} V_i^{ex}(\boldsymbol{r}),\ \ V_i^{ex}(\boldsymbol{r}) f(\boldsymbol{r}) = u_i(\boldsymbol{r}) \int d\boldsymbol{r}'\, \frac{u_i(\boldsymbol{r}')^* f(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|}, \tag{21.12}$$

where the summation $\sum_i^{\uparrow\uparrow}$ runs over the spin-orbitals with the same spin as i. For an atom with atomic number Z, the Fock operator

$$\hat{f} = \frac{p^2}{2m} - \frac{Z}{|\boldsymbol{r}|} + V^d(\boldsymbol{r}) - V^{ex}(\boldsymbol{r}) \tag{21.13}$$

allows us to write the HF equations in the one-body form

$$\hat{f}\ u_i(\boldsymbol{r}) = \varepsilon_i\ u_i(\boldsymbol{r}). \tag{21.14}$$

This form looks simple, but the problem is highly nonlinear, and the solution is usually sought numerically by iteration. The HF equations have a complete orthonormal set of solutions. The lowest N spin-orbitals allow us to build the determinantal wave functions. The rest are called *virtual orbitals*; they are not directly related to any experiment involving excited states, but they are often useful for generating multi-determinantal developments, like the Configuration Interaction expansion. The presence of the exchange term is essential in order to explain, for instance, the covalent bond and the cohesive energy of metals. A conduction electron feels the attraction of a piece of metal, which is neutral, because of the exchange contribution, which physically means that in the metal an electron manages to keep the electrons of the same spin far enough away to produce a sort of *Fermi hole* where the positive charge prevails.

This is just the beginning of the many-body theory. The typical error in the energy of valence electrons in the HF method is on the order of 1 eV, which may be relatively small, compared to the atomic ground state energies, but is comparable to the ionization potentials and to the energy of a typical chemical bond, so it is a serious error. In addition, particularly in the case of heavy elements, the effects of relativity on the inner electrons and also indirectly on the valence electrons are so strong that one must start with the relativistic formulation.[2]

[2] In the relativistic Dirac's theory, one manages to set up a formulation (Dirac–Fock approximation) despite the absence of a variational principle like the one that holds in the Schrödinger theory.

Accurate results are finally obtained my mixing the Hartree–Fock (or Dirac–Fock) determinant with many orthogonal excited determinants. This is configuration mixing, which is accurate, however, a computationally demanding, brute force approach, prohibitive for large systems.

Green's function methods and many-body perturbation theories, as well as Density Functional methods, are used for that.

21.4 The Density Functional (DF) Method

Despite the enormous increase of computational power, it is very important to devise cheap but accurate methods in order to broaden the scope of our calculations. W. Kohn deserved a Nobel prize by inventing, with his coworkers, a powerful and practical method, which then opened the way to much further progress.

The Hartree–Fock method includes the external potential V_{ext}, the Coulomb term (21.10) and exchange interactions via equation (21.11), but is not quite accurate. We may say that the self-consistent potential is missing a contribution, which is called correlation potential and is not known. Hohenberg–Kohn theorem[3] states that the density $n(\overrightarrow{r})$ of the electrons in the ground state of an N-electron liquid determines the ground state energy E_0 and also the ground state wave function ψ_0. Symbolically, we may write

$$n \Rightarrow E_0, \psi_0.$$

In other terms, the ground state energy and wave function is a universal functional of the ground state density. Moreover, one cannot modify V_{ext} without also changing n. Thus,

$$n \Rightarrow V_{ext} \Rightarrow E_0, \psi_0,$$

although the last implications require solving the Schrödinger equation. Indeed, we can solve it approximately, provided that we can find a good approximation for V_{xc}. Consider the energy functional

$$E_0[n] = T[n] + \frac{1}{2}\int dx dx' \frac{n(x)n(x')}{|x - x'|} + \int dx\, V_{ext}(x)n(x) + E_{xc}[n], \quad (21.15)$$

where $E_{xc}[n]$ is called the exchange-correlation functional. It is clear that all the problem rests with the unknown $E_{xc}[n]$, and any approximation for $E_{xc}[n]$ leads to a scheme for calculating the ground state energy and density by looking for the $n(x)$ which minimizes the energy with the given number of electrons.

Kohn and Sham[4] noted that everything except the last term is just the Hartree-Fock functional (21.9), provided that n is the actual density of the system under

[3] Hohenberg and Kohn (1964).

[4] Kohn and Sham (1965).

study; for the last term, one can use approximations drawn from the theory of the Fermi liquid, like

$$E_{xc}[n] = -\frac{3}{2}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int n^{\frac{4}{3}} d^3r - 0.056 \int \frac{n^{\frac{4}{3}}}{0.079 + n^{\frac{1}{3}}} d^3r. \tag{21.16}$$

So, one must solve a problem which is not more demanding than a Hartree–Fock calculation, but including correlation effects approximately via the last term. In actual practice, various $E_{xc}[n]$ have been tried, but overall the method has been highly successful. In problems involving magnetism and currents, one needs to extend the formalism. The so-called "current density" functional formalism has been proposed by Vignale and Rasolt.[5] Another very important extension was obtained by Runge and Gross, who formulated the time-dependent density functional formalism.[6]

References

Hohenberg, P., Kohn, W.: Phys. Rev. **136**, B864 (1964)
Kohn, W., Sham, L.J.: Phys. Rev. **140**, A1133 (1965)
Runge, E., Gross, E.K.U.: Phys. Rev. Lett. **52**, 997 (1984)
Vignale, G., Rasolt, M.: Phys. Rev. Lett. **59**, 2360 (1987)

[5] Vignale and Mark Rasolt (1987).

[6] Runge and Gross (1984).

Chapter 22
Discrete Models

Many problems are easier to grasp if we can make them discrete.

22.1 Matrices and Useful Models

In some problems the states that mix in a significant way are few; then, the method of matrices lends itself to insightful (if qualitative) descriptions. The simplest model is 2×2:

$$H = \begin{pmatrix} E_a & V, \\ V^* & E_b \end{pmatrix}.$$

This has many applications: diatomic molecules, etc. For example, if E_a, E_b are atomic levels and V a mixing term, which may represent a covalent bond, you get a lower *bonding* level and an *anti-bonding* one. This is a model for a molecule such as H_2^+. This theoretical model is rough, since it accounts for the energy gain of the electron which is delocalized but ignores interaction and correlation effects; however, qualitatively it works, while no explanation of the covalent bond is classically possible. Matrix models like this, with diagonal elements standing for atomic orbital energies and off-diagonal elements, represent electron hopping matrix elements, have been widely used for a rough description of molecular orbitals in polyatomic molecules, especially in organic chemistry, and are called Linear Combination of Atomic Orbitals (LCAO). For their great simplicity and their ability to explain some experimental trends, have quite a favorable cost-benefit ratio.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_22

As a second example, we consider neutrino oscillations. Neutrinos are difficult to detect and therefore require large detectors. Electron neutrinos can use the weak reaction

$$\nu_e + {}^{71}Ga \rightarrow {}^{71}Ge + e^- \tag{22.1}$$

where e^- is an electron. The problem originated in the '60's when it was found that the solar electron neutrinos are definitely fewer than the expectations based on the energy output. Neutrino oscillations have since been observed using reactors and accelerator experiments. There are 3 neutrino flavors, and we denote their spinors by $|\nu_\alpha\rangle$, with $\alpha = 1, 2$ and 3, respectively, for electron, muon and tauon neutrinos. The Hamiltonian is not diagonal on this basis, but on the basis $|\nu_i\rangle$ of mass eigenstates, with i = 1, 2, 3. On this basis, $H = \text{diag}(m_1, m_2, m_3)$, where the masses m_i are obviously relativistic (the rest masses are a fraction of 1 eV, while the observed energies are several Mev). The energy eigenstates (also called mass eigenstates) for the neutrinos ν_i, i = 1, 2, 3 are related to the flavor eigenstates by a 3×3 unitary matrix called the Pontecorvo, Maki, Nagakawa, Sakada matrix $U_{\alpha i}$. When only two flavors are important, this unitary matrix depends only on one parameter θ and may be written as

$$U = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}.$$

The probability that a particle created at time $t = 0$ as a flavor α neutrino is detected at time t by a detector for flavor β is readily found to be $P = |\langle \nu_\beta(t)|\nu_\alpha\rangle|^2$. This is a simple, phenomenological, but working, model of neutrino oscillations.

22.2 Gauge Changes in Lattice Models

We know from Chap. 15 (Sect. 15.1.5) that in Quantum Mechanics one can change the phase of a wave function arbitrarily because the change corresponds to a different choice of the gauge, involving the scalar potential V and the vector potential $\boldsymbol{A}$. In models with charged particles on discrete lattices, with sites connected by hopping Hamiltonian matrix element τ, this basic property holds like in continuous descriptions. For example, consider a *mini-lattice* of just two sites, with site energies 0 and ϵ:

$$H = \begin{pmatrix} 0 & \tau \\ \tau & \epsilon \end{pmatrix}. \tag{22.2}$$

In order to mimic the effect of an electric field $\boldsymbol{E}$ across a bond of the lattice, one can introduce an energy shift across the bond connecting two sites most simply by introducing a scalar potential difference $\Delta\epsilon = \hbar\omega$ between the sites connected by τ:

$$\begin{pmatrix} 0 & \tau \\ \tau & \epsilon \end{pmatrix} \rightarrow H_V = \begin{pmatrix} \frac{\hbar\omega}{2} & \tau \\ \tau & \epsilon - \frac{\hbar\omega}{2} \end{pmatrix}. \tag{22.3}$$

Alternatively, one can describe the same physics by introducing a time-dependent vector potential $\overrightarrow{A}(t)$ at the bond by a rule known as the the Peierls prescription:

$$\tau \to \tau \exp\left[\frac{2\pi i}{\phi_0}\int_{bond} \overrightarrow{A}.d\overrightarrow{r}\right], \tag{22.4}$$

where $\phi_0 = \frac{hc}{e}$, about 4×10^{-7} Gauss cm^2 is the fluxon, or flux quantum, the integral extends over the oriented bond and $\frac{2\pi}{\phi_0}\int_{bond} \overrightarrow{A}.d\overrightarrow{r} = \omega t$, a pure number.

The vector potential is related to the field by $\overrightarrow{A} = c\overrightarrow{E}t$. So, in this gauge, we set $\omega = \frac{2\pi}{\phi_0}\int_{\text{bond}} \boldsymbol{A}.d\boldsymbol{r}$, where $\phi_0 = \frac{hc}{e}$ is the flux quantum. The Hamiltonian is

$$H \to H_A(t) = \begin{pmatrix} 0 & \tau e^{i\omega t} \\ \tau e^{-i\omega t} & \epsilon \end{pmatrix}. \tag{22.5}$$

Since the time dependence is due to the vector potential, I append a label A to the amplitudes calling them a_A and b_A; they are found by solving the time-dependent S.E.; one finds

$$\begin{cases} i\dot{a_A} = \tau b_A(t)e^{i\omega t} \\ i\dot{b_A} = \epsilon b_A(t) + \tau a_A(t)e^{-i\omega t}. \end{cases} \tag{22.6}$$

The amplitudes obtained from H_V will be a_V and b_V. The gauge transformation is readily seen to be

$$a_V = a_A(t)\exp\left[\frac{-i\omega t}{2}\right], \tag{22.7}$$

$$b_V = b_A \exp\left[\frac{i\omega t}{2}\right]. \tag{22.8}$$

22.3 Flux Quantization

A neat example shows the flux quantization. The model Hamiltonian

$$H = \begin{pmatrix} 0 & \tau_{12} & 0 \\ \tau_{21} & 0 & \tau_{23} \\ 0 & \tau_{32} & 0 \end{pmatrix} \tag{22.9}$$

with $\tau_{12} = \tau_{23} = 1$ may represent the molecular orbitals of a symmetric triangular molecule. If we now set $\tau_{12} = \tau_{21}^* = e^{i\gamma}$ with real γ, the complex bond bears a vector potential, and consequently, the model is pierced by a magnetic flux. The ground state energy of the model is a periodic function of the flux, and the period is a flux quantum (Fig. 22.1).

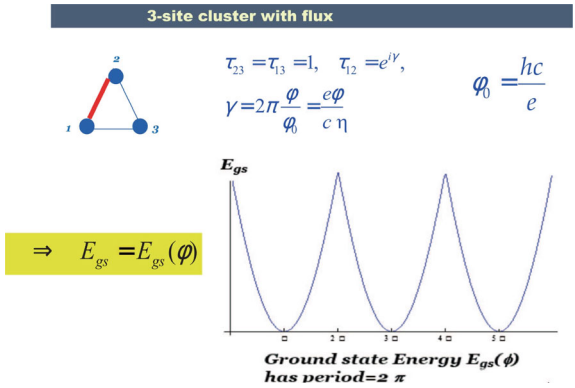


Fig. 22.1 The ground state energy of the triangular model as a function of γ

22.4 Infinite Discrete Models and the Local Density of States

The infinite chain is a prototype model of an infinite one-dimensional solid. The sites of the chain are labeled by integers, and the Hamiltonian describes nearest-neighbor hopping. In second quantization, it has the form:

$$H = \tau \sum_n c_n^\dagger (c_{n+1} + c_{n-1}). \tag{22.10}$$

The first-quantized form is

$$H|n\rangle = \tau \sum_n (|n+1\rangle + |n-1\rangle). \tag{22.11}$$

Expanding the wave functions on the basis of the sites, $|\psi\rangle = \sum_{n=-\infty}^{\infty} \alpha_n |n\rangle$, one finds for the eigenstate of energy ϵ_q the S.E.

$$H|\psi\rangle = \sum_{n=-\infty}^{\infty} \alpha_n H|n\rangle = \epsilon_q |\psi\rangle, \tag{22.12}$$

which leads to the recursion relation $\epsilon_q \psi_n = \tau(\alpha_{n+1} + \alpha_{n-1})$. This is solved by

$$\epsilon_q = 2\tau \cos(q)$$

and

$$\alpha_n = \frac{e^{iqn}}{\sqrt{N}}.$$

The distribution of the energy eigenfunctions is obviously of interest. The local density of states (LDOS) defined by

$$\rho(\omega) = \sum_k |\langle 0|k\rangle|^2 \delta(\omega - \epsilon_k) \tag{22.13}$$

gives the probability of finding the particle at site 0 with energy $\hbar\omega$ (but it is the same on all sites in this model). $\rho(\omega)$ contains much information and allows us to visualize the continuum. In one dimension, let us call it $\rho(\omega)_1$. Using the fact that $\sum_k = \frac{1}{2\pi}\int_{-\frac{\pi}{a}}^{\frac{\pi}{a}} dk$, one easily does the integral with the δ and finds that the density of states vanishes outside the band $-2t_h \leq \omega \leq 2t_h$, and within the band,

$$\rho(\omega)_1 = \frac{1}{\pi\sqrt{4t_h^2 - \omega^2}}. \tag{22.14}$$

This is shown in Fig. 22.2 left. The divergence at the band edges is topological, i.e., it is common to all the one-dimensional models. Indeed, the group velocity of the particle goes to zero there, and so it is likely that the particle is found nearby.

The local density of states is also easily obtained as an integral in two and three dimensions. In two-dimensions, one finds that

$$\rho_2(\omega) = \frac{1}{2\pi}\int_{-\pi}^{\pi} dk \rho_1(\omega - 2t_h\cos(k)). \tag{22.15}$$

This is shown in Fig. 22.2 center. The band edge singularity has become a finite jump, while a divergence of the derivative at $\omega = 0$ is an example of the so-called Van Hove singularities. In Solid State Physics, such singularities play a role in the analysis of the optical absorption spectra. In a similar way, one finds for the three-dimensional case,

$$\rho_3(\omega) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \rho_2(\omega - 2t_h\cos(k)). \tag{22.16}$$

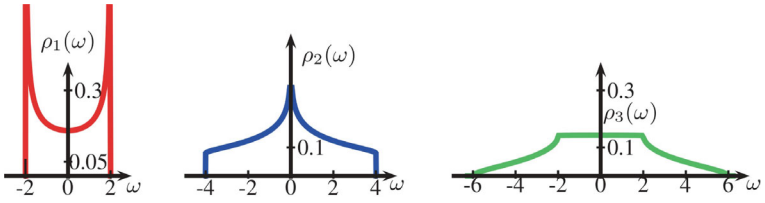


Fig. 22.2 Density of states of the one-dimensional tight-binding chain (left), square lattice (center) and cubic lattice (right). Note the Van Hove singularities at the band edges and, in the three-dimensional case, at $\omega = \pm 2$

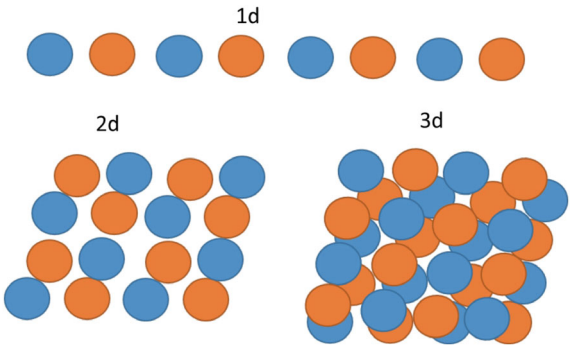


Fig. 22.3 Slices of Bipartite graphs in d, two-dimensional and three-dimensional

This is shown in Fig. 22.2 right. The band edge singularities at $\omega = \pm\omega_{edge}$ with $\omega_{edge} = 6t_h$ are of the square root type, and this is typical of three-dimensionality.

The densities of states are all symmetric around $\omega = 0$. This is readily understood, since all three lattices are bipartite, that is, they have the property that all sites can be painted red or blue, with every red site having only blue nearest neighbors and every blue site having only red nearest neighbors. This implies that a change $t_h \to -t_h$ is just a gauge change, equivalent to changing the sign of all the red sites; no physical observable can be affected. On the other hand the Hamiltonian changes sign. Therefore, the spectrum is symmetric for $\omega \to -\omega$, and $t_h \to -t_h$ brings an eigenfunction to an eigenfunction corresponding to an opposite eigenvalue (Fig. 22.3).

22.5 Exactly Solved Non-periodic Infinite Model

The above models are easily solved thanks to the symmetry. A few more models that can be solved, even for large or infinite systems, are naturally of interest. We shall find the following simple model very enlightening. Consider the Hamiltonian

$$H = \begin{pmatrix} \epsilon_1 & g & g & \cdots \\ g & \epsilon_2 & g & \cdots \\ g & g & \epsilon_3 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{22.17}$$

For the moment, we may take it as a model with site energies ϵ_n, where the electron can hop from one site to any other with the same amplitude g. This model can be solved for any size of the system. Let v_n denote the nth component of any eigenvector. The S.E. gives us

$$\epsilon_1 v_1 + g \sum_{j \neq 1} v_j = E v_1, \tag{22.18}$$

where E is the eigenvalue. Setting $S = \sum_j v_j$, we obtain $v_1 = \frac{gS}{E+g-\epsilon_1}$. Since the equation for the n-th component is of the same form, namely $v_n = \frac{gS}{E+g-\epsilon_n}$, we readily arrive at the eigenvalue equation

$$\frac{1}{g} = \sum_n \frac{1}{E + g - \epsilon_n} \tag{22.19}$$

and the problem is reduced to the solution of a transcendental equation. This solution is versatile. It will be used in Sect. 25.6.1 to solve the Richardson model of superconductivity.

22.6 The Current Operator on a Discrete Basis

Consider the following prototype second-quantized model for a quantum wire:

$$H = t_h \sum_i (c^\dagger_{i+1} c_i + h.c.); \tag{22.20}$$

the corresponding expression in first quantization would read as

$$H = t_h \sum_i (|i+1\rangle\langle i| + h.c.)$$

and would describe the hopping of one electron. Instead, (22.20) is a many-body Hamiltonian that represents electrons doing the same but obeying the anti-commutation relations.

The continuum representation of the current density (Eq. (10.20)) cannot apply, but we can build a new one still based on the continuity equation; the difference is that now the density of spin σ at site i is represented by

$$\rho(i, \sigma) = e\hat{n}_{i\sigma} = c^\dagger_{i\sigma} c_{i\sigma}.$$

The spin index can be understood when it does not play a role. In writing the continuity equation for the current J for this discrete model, we may set

$$\frac{dJ}{dx} \to J_m - J_{m-1}$$

and so

$$\frac{\partial \rho}{\partial t} = J_m - J_{m-1}.$$

The time derivative is obtained from the Heisenberg equation of motion. In this way we arrive at the current operator

$$J_m = \frac{et_h}{i\hbar}(c^\dagger_{m-1}c_m - c^\dagger_m c_{m-1}). \tag{22.21}$$

An expression like $J_m = \frac{et_h}{i\hbar}(c^\dagger_m c_{m+1} - c^\dagger_{m+1}c_m)$ is equally correct since it is physically equivalent. In Chap. 25 we shall see applications of this result.

Chapter 23
Pancharatnam Phase and Berry Phase

Parametric Hamiltonians convey the indirect influence of the rest of the Universe on the system under study, and this influence is formally similar to a magnetic field acting in abstract parameter space.

23.1 The Pancharatnam Phase

The Indian physicist S. Pancharatnam, working on quantum Optics in 1956, introduced[1] the novel concept of a geometrical phase. Let $H(\xi)$ denote a Hamiltonian that depends on one external parameter ξ, like an angle or a length, and its ground state $|\psi(\xi)\rangle$. One can define a phase difference $\Delta\varphi_{12}$ between two ground states $|\psi(\xi_1)\rangle$ and $|\psi(\xi_2)\rangle$ and obtain by calculating both by the same algorithm or computer code, just with a change in the parameter:

$$\langle\psi(\xi_1)|\psi(\xi_2)\rangle = |\langle\psi(\xi_1)|\psi(\xi_2)\rangle|e^{-i\,\Delta\gamma_{12}}. \tag{23.1}$$

(Of course, I am assuming that the two ground states are not orthogonal.) The phase difference $\Delta\varphi_{12}$ obtained in this way as a byproduct of the calculation cannot have any physical meaning, since the phase of any quantum wave function is arbitrary and moreover, for a charged particle, it depends on the choice of a gauge. But wait. Now consider 3 points in parameter space and calculate the total phase γ in $\xi_1 \to \xi_2 \to \xi_3 \to \xi_1$. Now, what matters is the Pancharatnam phase,

$$\gamma = \Delta\gamma_{12} + \Delta\gamma_{23} + \Delta\gamma_{31}. \tag{23.2}$$

[1] Pancharatnam (1956).

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_23

The phase of each ψ can be changed at will by a gauge transformation, but strikingly, all these arbitrary changes cancel in the calculation of γ. Actually, γ is gauge-independent! This result holds in any closed circuit defined by more than 2 points. There is no excuse for declaring γ a physically irrelevant quantity. It can be an observable. *Some observables are not eigenvalues of Hermitian operators.* This is striking, because it implies that the abstract parameter space is indeed a hidden part of the reality.

In order to see how a combination of three (or more) arbitrary phases can be physically relevant, one can use a discrete model for a triatomic molecule with atomic orbitals $|\psi_a\rangle$, $|\psi_b\rangle$, $|\psi_c\rangle$; the Hamiltonian H has matrix elements connecting them. Let τ_{ab} be the matrix element connecting sites a and b, which allows electron hopping. This is just the LCAO model of a molecule (see Sect. 22.1). Now, following the Peierls prescription, we can *switch on* a vector potential $\overrightarrow{A}$ with

$$\tau_{ab} \rightarrow \tau_{ab} e^{\frac{2\pi i}{\phi_0} \int_a^b d\overrightarrow{r} \cdot \overrightarrow{A}}, \tag{23.3}$$

where $\phi_0 = \frac{hc}{e} = 4 \times 10^{-7}$Gauss cm^2 is the flux quantum or *fluxon*. In the case of a biatomic molecule like H_2, this change has no consequences, but with 3 or more atoms, the physical meaning is that a magnetic flux ϕ pierces the molecule. Changing ϕ modifies all the energy levels, except that changing ϕ by a fluxon is a gauge change (see Chap. 22).

Thus, the Pancharatnam phase γ can originate in a subtle way from the parameters in the Hamiltonian that stand for the influence of the external world.

23.2 The Berry Phase

Next we wish to make two changes: we wish to go from three points to infinitely many (continuum limit) and have the phase vary with continuity. Starting from an abstract parameter space, in which the parameters can be chosen in many ways, we wish to define useful measurable quantities which characterize the system in a nontrivial way. To this end, we consider a model Hamiltonian which depends on several magnitudes, like lengths, angles and field strengths. For the sake of argument, let us assume that it depends on two (or possibly three) such quantities (see Fig. 23.1). We can consider a two (or possibly three)-dimensional parameter space, where each point $\overrightarrow{R}$ corresponds to some Hamiltonian with some phase differences between sites.

What does the Pancharatnam phase become?

To understand this question, consider a large number N of states $m = 1, 2, \cdots N$, forming a finely spaced necklace (Fig. 23.1). We can afford to be general, so let us assume that the system is in the n-th eigenstate of the Hamiltonian with the parameters corresponding to $m = 1$. In order to avoid complications, we also assume a non-degenerate spectrum. If the wave functions are computed on the same footing, their

phases will also behave smoothly in going from point 1 to 2; if ψ denotes the wave function at 1, we expect to find in 2 the function $(\psi + \delta\psi)e^{-i\delta\gamma_{12}}$, where $\delta\gamma_{12}$ is the phase change. In other terms,

$$\langle\psi_1|\psi_2\rangle = \langle\psi|\psi + \delta\psi\rangle e^{i\delta\gamma_{12}}; \tag{23.4}$$

but we know from Chap. 20 that the first-order change $\delta\psi$ must be orthogonal to ψ. Thus,

$$\langle\psi_1|\psi_2\rangle \sim e^{-i\delta\gamma_{12}} \sim 1 - i\delta\gamma_{12}.$$

On the other hand, it is natural to go to the continuum limit; to first order,

$$\psi_2 = \psi_1 + \nabla\psi.d\boldsymbol{r}, \tag{23.5}$$

where ∇ stands for the gradient operator in parameter space.

Therefore, comparing with (23.4), we conclude that

$$\delta\gamma_{12} = i\langle\psi|\nabla\psi\rangle.$$

In the continuum limit, the gauge invariant phase collected in the necklace

$$\gamma = i\oint\langle\psi|\boldsymbol{\nabla}\psi\rangle d\boldsymbol{R} \tag{23.6}$$

is the *Berry phase* while the so-called *Berry connection*

$$\boldsymbol{A}_n = \langle\psi_n|\boldsymbol{\nabla}\psi_n\rangle \tag{23.7}$$

is gauge-dependent and can be thought of as a sort of vector potential. But, in analogy with electromagnetism,

$$curl\,\boldsymbol{A}_n = \boldsymbol{B}_n \tag{23.8}$$

is gauge-independent like a magnetic field and is observable. In the present context, $\boldsymbol{B}_n$ takes the name of curvature; some curvature must create a flux through the yellow region of Fig. 23.1 if the Berry phase is different from 0. The term "curvature" is also used for the antisymmetric matrix

$$\Omega_{\alpha\beta} = -2Im\langle\partial_\alpha\psi|\partial_\beta\psi\rangle. \tag{23.9}$$

Then, the Berry phase can be rewritten as

$$\gamma = \frac{1}{2}\int_\gamma d\xi^\alpha \wedge d\xi^\beta\Omega_{\alpha\beta}(\xi), \tag{23.10}$$

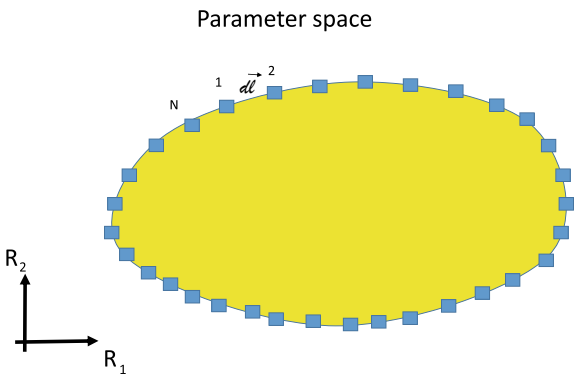


Fig. 23.1 Necklace of states in a two-dimensional parameter space

where $\wedge$ is the antisymmetrized product such that $\vec{a} \wedge \vec{b} = -\vec{b} \wedge \vec{a}$. The study of the behavior of the system in parameter space turns out to be very useful to explain the properties in real space, too. The higher the dimensionality of the parameter space, the higher is the advantage it can provide.

The Berry phase,[2] which was introduced in 1983, is very fashionable for the important developments in various fields of physics. For instance, the modern theory of the polarization of solids is based on the use of the Berry phase.[3]

23.3 Chern Number

Some results from topology are used in the present context and make the study of the parameter space quite relevant. The Gauss–Bonnet–Chern theorem states that for any closed surface S^2 in 3 dimensions, with $\boldsymbol{A}$ a Berry connection,

$$\int_{S^2} curl\boldsymbol{A} \cdot \hat{n} d\sigma = 2\pi C_1, \tag{23.11}$$

where C_1 is an integer which is called Chern number. The fact that an integer appears in a calculation with no integers in input is quite striking and is very important for the physical applications. There are important applications of the Berry phase in Quantum Transport. These are postponed to Chap. 26.

[2] Berry (1984).

[3] See, e.g., Resta (2000).

23.4 Molecular Aharonov–Bohm Effect

In molecular studies one can often simplify matters by adopting the Born–Oppenheimer approximation, which consists of the neglect of nuclear momenta. One then remains with the electronic problem. Often the molecular energy surface as a function of the nuclear coordinates has several equivalent minima, and the nuclear degrees of freedom can hardly tunnel between them. Then, the kinetic energy of the nuclei does not play a role, and one can observe a static JT effect with broken symmetry. A popular approach consists in considering several equivalent minima and then letting them interact. Group theory is generally very helpful in finding the normal modes of vibration of the molecule and the potentials whereby the modes interact with the electronic wave functions.

A very interesting particular example was discovered long ago[4] and is known as the molecular Aharonov–Bohm effect. For our purposes, the approximate molecular wave function can be taken in the form

$$\Phi = \psi_{el}(\xi, x)\psi_{nuc}(\xi), \tag{23.12}$$

where ψ_{el} and ψ_{nuc} are electron and nuclear wave functions and x, ξ electron and nuclear coordinates. Such wave functions do not really exist in principle, since electrons and nuclei are entangled (see Chap. 27), but they are useful in approximate treatments. Now suppose that we want to study the roto-vibrational spectrum of the molecule. To this end, we must determine the slow nuclear motion, averaged over the fast electronic degrees of freedom. Let $\langle \cdots \rangle_{el}$ denote the latter average and $p_\alpha = -i\hbar\frac{\partial}{\partial \xi_\alpha}$ denote the operator of the α component of the nuclear momentum. The *effective* nuclear momentum component π_α acting on the nuclear wave function $\psi_{nucl}(\xi)$ must be obtained by averaging over the "fast" electronic wave functions. So, it has components

$$\pi_\alpha \psi_{nucl} = \langle \psi_{el} | p_\alpha | \Phi \rangle_{el} = -i\hbar \frac{\partial}{\partial \xi_\alpha} \psi_{nucl}(\xi) + \langle \psi_{el}(\xi, x) | \frac{\partial \psi_{el}(\xi, x)}{\partial \xi_\alpha} \rangle_{el} \psi_{nucl}(\xi). \tag{23.13}$$

Trivial Case: Vanishing Berry Connection

Now,

$$A_\alpha = \langle \psi_{el} | p_\alpha \psi_{el} \rangle = -i\hbar \frac{\partial}{\partial \xi_\alpha} \psi_{nucl}(\xi) + \langle \psi_{el}(\xi, x) | \frac{\partial \psi_{el}(\xi, x)}{\partial \xi_\alpha} \psi_{el}(\xi, x) \rangle \tag{23.14}$$

looks very much like a vector potential and is actually a Berry connection. If the electron wave functions can be taken as real, the Berry connection vanishes, since $A_\alpha = \frac{\partial}{\partial \xi_\alpha} \int dx |\psi_{el}(\xi, x)|^2$, and the electron wave functions are normalized to 1. Then, from the nuclear wave function one can derive a rotational and vibrational spectrum. The vibrational spectrum is well approximated by a collection of harmonic

[4] Longuet-Higgins et al. (1958).

terms. The rotational spectrum is that of a rotator, and we saw in Sect. 13.5 that the eigenvalues are labeled by integer quantum numbers.

Non-vanishing Berry Connection in $e \times \epsilon$ Symmetry

But there is an interesting complication in some geometries, like the so-called $e \times \epsilon$ symmetry, when both electron and nuclear wave functions transform under the symmetry operations like the pair (x, y). This case requires the use of Group Theory and is discussed in detail elsewhere.[5] One can show using the symmetry that we can cast the Hamiltonian in the form

$$H = -\frac{\hbar^2}{2M_0}\left(\frac{\partial^2}{\partial q_x^2} + \frac{\partial^2}{\partial q_y^2}\right) + \lambda(q_x\sigma_x + q_y\sigma_y) + Kq^2, \tag{23.15}$$

where M_0 is a large mass (but we shall consider it infinite), σ_x and σ_y are Pauli matrices and K is a constant. The limit of large M_0 means physically that we are looking for a frozen nuclear displacement. In terms of angular displacements, one can write

$$H = \lambda q \begin{pmatrix} \sin(\theta) & \cos(\theta) \\ \cos(\theta) & -\sin(\theta) \end{pmatrix}, \tag{23.16}$$

and the eigenvalues are $\lambda q + Kq^2$ and $-\lambda q + Kq^2$; both eigenvalues are independent of θ. Hence, the potential energy surfaces $E(q)$ are obtained by rotating two intersecting parabolas around the energy axis, and both parabolas have minima away from $q = 0$. The eigenvectors are

$$\chi_-(\theta) = \begin{pmatrix} \cos(\theta/2 + \pi/4) \\ -\sin(\theta/2 + \pi/4) \end{pmatrix}, \tag{23.17}$$

and

$$\chi_+(\theta) = \begin{pmatrix} \sin(\theta/2 + \pi/4) \\ \cos(\theta/2 + \pi/4) \end{pmatrix}. \tag{23.18}$$

By symmetry,the pair (x, y) is doubly degenerate. It turns out that if the electron wave functions are taken to be real, then the nuclear wave functions transform as the components of a spin 1/2 and must change sign under a 2π rotation. This is absolutely not acceptable, since the nuclear wave functions must be single-valued! We did an honest calculation. Why did we get a wrong result? We can fix it, but must insert phase factors:

$$\chi_-(\theta) = e^{i\theta/2} \begin{pmatrix} \cos(\theta/2 + \pi/4) \\ -\sin(\theta/2 + \pi/4) \end{pmatrix}, \tag{23.19}$$

[5] See, for instance, my book "Topics and Methods in Condensed Matter Theory", Springer Verlag, Sect. 9.6.2 (page 196).

and

$$\chi_+(\theta) = e^{-i\theta/2} \begin{pmatrix} \sin(\theta/2 + \pi/4) \\ \cos(\theta/2 + \pi/4) \end{pmatrix}. \tag{23.20}$$

This Berry phase compensates for the sign change in a 2π rotation.

23.5 Su–Schrieffer–Heeger Model: Berry Connection and Winding Number

The Berry connection for this model is

$$A_-(k) = i\langle k - |\frac{d}{dk}|k-\rangle = -\frac{1}{2}\frac{d\phi_k}{dk}; \tag{23.21}$$

one can verify that the winding number, defined as the integral of the Berry connection over the Brillouin Zone

$$g = -\frac{1}{\pi}\int_{BZ} A_-(k)dk, \tag{23.22}$$

vanishes if $v > w$ and is 1 if $v < w$. This is remarkable because changes in v and w do not change the result. The difference between the two insulating phases is topological. One cannot go from one insulating phase to the other (by changing the parameters) without a topological phase transition, which involves closing the gap.

References

Berry, M.V.: Proc. R. Soc. Lond. A **392**, 45 (1984)
Longuet-Higgins, H.C., Opik, U., Pryce, M.H.L., Sack, R.A.: Proc. R. Soc. A **244**, 1 (1958)
Pancharatnam, S.: Proc. Indian Acad. Sci. A **44**, 247 (1956)
Resta, R.: J. Phys.: Condens. Matter **12**, R107 (2000)

Chapter 24
Fano Resonances

Quantum Mechanics replaces smooth trends of Classical physics by sharp leaps, but the discrete energy levels are always resonant. This means that we have missed something general and important up to now.

24.1 The Dynamics of the Excited States

All the solvable models that we have seen in previous chapters lead to continuum spectra for unbound particles and discrete spectra for bound states; for example, the H atom gave us an infinity of discrete states and the electron–proton continuum. While this may be a useful first approach to the Physics of bound states, in reality it is only the beginning of the story, since all the excited states, once created, must evolve in some way. For example, the 2p level of Hydrogen is discrete, but is degenerate with a system comprising the atom in the 1 s state + a photon, whose energy belongs to a continuum. Including the coupling between the atom and the radiation field, we obtain a finite lifetime and a width to all excited states, in accord with the uncertainty principle. But, interestingly, the excited levels do not simply broaden. They acquire a structure when coupled to continua. One can add many more examples of resonances to those enlisted in Sect. 11.6:

1. The Feshbach resonances of molecular Physics are peaks in the electron-molecule elastic cross section. For instance, in electron scattering with SF_6, there is a resonance when the electron momentum p is such that De Broglie wavelength $\lambda = \frac{h}{p}$ is close to the Sulfur–Fluorine distance. The big F^- ions make a sort of cage around S and when the electron wave function fits in the cage, a temporary

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_24

negative ion is formed. In other terms, it takes sometime before the electron escapes.
2. Phenomena of autoionization in atoms, like the Auger effect (see Chap. 26).
3. In particle physics, most particles have a finite life time and are resonances, like the Δ^{++} resonance discovered by Fermi in $\pi-$proton scattering.

The discrete spectra often arise when a partial Hamiltonian is solved, but are not exact eigenstates when a more complete Hamiltonian is considered. In Chap. 17, when dealing with the one-electron atom, we ignored the coupling to the transverse electromagnetic waves; because of this coupling, all the excited states are resonances.

In order to make progress, let us start with one discrete state that we denote by $|0\rangle$ and one non-degenerate continuum $\mathcal{C}$ of states that we denote by $|k\rangle$. The continuum states would be energy eigenstates with energy ϵ_k, in the absence of the mixing with $|0\rangle$; they have a density of states $\rho^c(\omega)$, that we shall regard as a known quantity. Hence,

$$N = \int_{E_1}^{E_2} d\omega \rho^c(\omega)$$

is the number of states in the interval $E_1 < \omega < E_2$; it diverges in the continuum limit. In order to convert the sum into an integral, we need a suitable measure, like

$$\sum_k \to \frac{\Omega}{(2\pi)^3} \int d^3k,$$

where Ω is the volume of the system. In the end, all physical quantities have a finite value when $\Omega \to \infty$.

The discrete state would have a sharp energy ϵ_0 in the absence of the mixing; ϵ_0 is supposed to be known from the start. Some kind of coupling (denoted by $\hat{V}$) changes the discrete delta-like level into a *virtual level*; this has a width and can have structure. The λ set is complete, hence $\int_{-\infty}^{\infty} d\omega \rho_0(\omega) = 1$.
This is the projected density of states

$$\rho_0(\omega) = \sum_\lambda |\langle 0|\lambda\rangle|^2 \delta(\omega - \epsilon_\lambda), \tag{24.1}$$

where $H|\lambda\rangle = \epsilon_\lambda|\lambda\rangle$; the eigenstates $|\lambda\rangle$ are unknown (for the moment) and the eigenvalues ϵ_λ can be continuous or discrete. We already encountered this function in Chap. 22. The unperturbed quantity is $\rho_0^{(0)}(\omega) = \delta(\omega - \epsilon_0)$.

The projected density of states is built in into the resolvent, or Green operator

$$G(\omega) = (\omega - H + i\delta)^{-1}, \tag{24.2}$$

where δ stands for a positive infinitesimal, or $\delta = +0$ for short. $G(\omega)$ is the Fourier transform of the causal (i.e., proportional to $\theta(t)$) operator

$$G(t) = e^{-iHt}\theta(t).$$

It is causal and analytical in the upper half plane. Now, inserting a complete set, one finds

$$G_{00}(\omega) \equiv \langle 0|G(\omega)|0\rangle = \sum_{\lambda} \frac{|\langle 0|\lambda\rangle|^2}{\omega - \epsilon_\lambda + i\delta}, \tag{24.3}$$

and the relation to the local DOS is

$$\rho_0(\omega) = -\frac{1}{\pi} ImG_{00}(\omega). \tag{24.4}$$

Here, it is understood that $\sum_\lambda \equiv \int_{\mathcal{C}} d\epsilon_\lambda + \sum_{\lambda\in\mathcal{D}}$, where $\mathcal{D}$ is the set of the discrete eigenvalues outside the set $\mathcal{C}$ of energies of the unperturbed continuum. While ρ_c is an input quantity of the problem, ρ_0 must be found, and at this point, we must specify a model Hamiltonian.

24.2 Fano Model

It follows from the above consideration that the model we want must have the structure of the Fano model,[1] that we have encountered in Sect. 19.5.1, Eq. (19.51). The model is often written in second-quantized form, where the operators are meant to represent electrons and are meant to anticommute; it reads

$$H = H_0 + H_h,\ H_0 = \sum_{k\in\mathcal{C},\sigma} \epsilon_{k,\sigma} n_{k,\sigma} + \sum_{\sigma} \epsilon_{0,\sigma} n_{0,\sigma} \tag{24.5}$$

while

$$H_h = \sum_{k,\sigma} \{V_k a^\dagger_{k\sigma} a_{0\sigma} + h.c.\}. \tag{24.6}$$

This model does not specify the mixing mechanism. Since the problem arises in particle Physics, in Atomic as well as in Solid State Physics, the most diverse mechanisms may be operating: the Fano model simply says that something eventually mixes the continuum states with the localized state. In condensed matter Physics $|0\sigma\rangle$ might represent an atomic spin-orbital and $|k\sigma\rangle$ a free-particle state; in nuclear Physics the model could represent radioactive decay, and so on.

This is a one-particle model, so it is exactly solvable and is very useful for a qualitative analysis of real situations. The spin actually plays no role in the simplest

[1] Fano (1961).

version of the model; the many-body extension creates no difficulty since the determinantal eigenstates of H_h simply are formed with the eigen-orbitals of the one-body model multiplied by spin functions. There are no interaction terms (involving two creation and two annihilation operators) in the Fano model. The assumption of a non-degenerate continuum can be removed later by a direct extension of the present treatment.[2]

A second-quantized Hamiltonian like H is diagonalized by finding new one-body states $|\lambda\rangle$ which are linear combinations of the solutions for $V_k = 0$, namely, $|0\rangle$ and $|k\rangle$; the eigenstates $|\lambda\rangle$ must be such that the transformed Hamiltonian has the form:

$$\tilde{H} = \sum_{\lambda} \epsilon_\lambda \hat{n}_\lambda \tag{24.7}$$

where ϵ_λ are the exact eigenvalues. The expansion is

$$a_\lambda = \langle\lambda|0\rangle a_0 + \sum_k \langle\lambda|k\rangle a_k. \tag{24.8}$$

The inverse expansion is

$$a_k = \sum_{\lambda} \langle k|\lambda\rangle a_\lambda, \qquad a_0 = \sum_{\lambda} \langle 0|\lambda\rangle a_\lambda. \tag{24.9}$$

Let us use *the equation of motion method*. The equation of motion

$$[a_\lambda, \tilde{H}]_- = \epsilon_\lambda a_\lambda \tag{24.10}$$

must be true also in the original basis. Substituting (24.8) and (24.5), one finds the infinite system

$$\begin{aligned} (\epsilon_\lambda - \epsilon_0)\langle\lambda|0\rangle - \sum_k \lambda|k\rangle V_{0k} = 0 \\ (\epsilon_\lambda - \epsilon_k)\langle\lambda|k\rangle - \langle\lambda|0\rangle V_{k0} = 0. \end{aligned} \tag{24.11}$$

However, $\langle\lambda|k\rangle$ must be obtained as a distribution; by setting

$$\langle\lambda|k\rangle = \langle\lambda|0\rangle V_{k0} \left[\frac{P}{\epsilon_\lambda - \epsilon_k} + Z(\epsilon_\lambda)\delta(\epsilon_\lambda - \epsilon_k) \right], \tag{24.12}$$

[2] Davis and Feldkamp (1977).

where Z is an unknown, we may substitute in the first of (24.11) which yields, setting $\epsilon_\lambda = \omega$,

$$Z(\omega) = \frac{\omega - \epsilon_0 - Re(\Sigma(\omega))}{Im(\Sigma(\omega))}, \tag{24.13}$$

where

$$\Sigma(\omega) = \sum_k \frac{|V_{0k}|^2}{\omega - \epsilon_k + i\delta} \tag{24.14}$$

is the self-energy. Σ is an analytic function of $z = \omega + i\delta$, outside the real axis. If $\mathcal{C} = \{a \leq \omega \leq b\}$, then there is a cut just below the axis, with a and b as branch points. We must still find $\langle\lambda|0\rangle$ and appreciate the meaning of the self-energy. This is best done in terms of the density of states, as follows.

24.3 Working in a Subspace: The Self-Energy Operator

In many problems, it is convenient to put the Hamiltonian in a block form,

$$H = \begin{pmatrix} H_{AA} & H_{AB} \\ H_{BA} & H_{BB} \end{pmatrix}, \tag{24.15}$$

where A and B are different subspaces of the Hilbert space; that is, the Hilbert space is $A \bigcup B$. In the absence of H_{AB} they denote, respectively, the low-energy and high-energy parts. Then, a low-energy phenomenon occurs essentially in the A subspace but is subject to indirect influences from B due to H_{AB}. The resolvent matrix G is defined by

$$(\omega - H + i\delta)G(\omega) = 1, \tag{24.16}$$

where $i\delta$ is a small imaginary part added to the frequency in order to avoid divergences on the real axis; however, the imaginary part can be understood. More explicitly,

$$\begin{pmatrix} \omega - H_{AA} & -H_{AB} \\ -H_{BA} & \omega - H_{BB} \end{pmatrix} \begin{pmatrix} G_{AA} & G_{AB} \\ G_{BA} & G_{BB} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{24.17}$$

This is readily solved. The solution in A subspace reads as

$$G_{AA} = (\omega - H_{AA} - H_{AB}(\omega - H_{BB})^{-1}H_{BA})^{-1}. \tag{24.18}$$

Here, $-H_{AB}(\omega - H_{BB})^{-1}H_{BA}$ is a matrix self-energy that conveys the effects of the B subspace on the resolvent in the A subspace. That is, one can solve the problem within subspace A with an effective Hamiltonian $H = H_{AA} + \Sigma_{AA}$ where we may

write symbolically $\Sigma_{AA} = H_{AB}\frac{1}{\omega - H_{BB}}H_{BA}$. Now we apply this result to the Fano model, setting

$$H_{AA} = \epsilon_0 \hat{n}_0 \equiv \epsilon_0 |0\rangle\langle 0|,\ H_{BB} = \sum_k \epsilon_k |k\rangle\langle k|,\ H_{AB} = \sum_k V_{0k} |0\rangle\langle 0|.$$

So,

$$H_{AB}\frac{1}{\omega - H_{BB}}H_{BA} = \sum_k |0\rangle\langle k| \frac{1}{\omega - \sum_{k'} |k'\rangle \epsilon_{k'} \langle k'|} \sum_{k"} |k"\rangle V_{k"0},$$

and finally, we obtain the exact Green's function

$$G_{00} = \frac{1}{\omega - \epsilon_0 - \Sigma(\omega)}, \tag{24.19}$$

where $\Sigma(\omega)$ is the *self-energy* (24.14). The other elements of the resolvent matrix can also be derived in this way.

This is a complex function of ω, in which $\sum_k$ stands for an integral;

$$\begin{aligned} \Sigma_1(\omega) \equiv \mathrm{Re}\Sigma(\omega) &= \mathrm{P}\sum_k \frac{|V_{k0}|^2}{\omega - \epsilon_k} \\ \Sigma_2(\omega) \equiv \mathrm{Im}\Sigma(\omega) &= -\pi \sum_k |V_{k0}|^2 \delta(\omega - \epsilon_k). \end{aligned} \tag{24.20}$$

Outside $\mathcal{C}$, $\Sigma_2 = 0$. Σ is a **Herglotz**[3] function, that is, $-\pi^{-1}\mathrm{Im}\Sigma(\omega) \geq 0$; it follows that $G_{00}(\omega)$ is also Herglotz, which is important to ensure that $\rho_0 \geq 0$.

In order to obtain new insight on the wave functions, we change variables in Eq. (24.3)

$$G_{00}(\omega) = \int d\epsilon_\lambda \rho_0(\epsilon_\lambda) \frac{|\langle 0|\lambda\rangle|^2}{\omega - \epsilon_\lambda + i\delta}, \tag{24.21}$$

where we used the one-to-one correspondence between states and energies, valid thanks to the assumption of a non-degenerate continuum. In this way, with λ_ω denoting the eigenstate at $\epsilon = \omega$,

$$\mathrm{Im}G_0(\omega) = -\pi \int d\epsilon_\lambda \rho(\epsilon_\lambda) |\langle 0|\lambda\rangle|^2 \delta(\omega - \epsilon_\lambda) = -\pi\rho(\omega)|\langle 0|\lambda_\omega\rangle|^2, \tag{24.22}$$

and so,

$$|\langle \lambda_\omega|0\rangle|^2 = \frac{-1}{\pi\rho(\omega)}\mathrm{Im}G_{00}(\omega) = \frac{\rho^0(\omega)}{\rho^c(\omega)}.$$

[3] Gustav Herglotz (1881–1953) was a German mathematician.

Since the phase can be chosen at will, we are free to set

$$\langle \lambda_\omega | 0 \rangle = \sqrt{\frac{\rho^0(\omega)}{\rho^c(\omega)}}. \tag{24.23}$$

Together with Eq. (24.12), this result yields all the amplitudes. Besides,

$$\Sigma_2(\omega) = -\pi \rho^0 V_\omega^2, \tag{24.24}$$

where, following Fano, I have set $V_\omega = V_{0k}$ with $\epsilon_k = \omega$.

This technique is useful when it is desirable to work explicitly in a subspace A of the Hilbert space (in this case represented by the discrete state) while taking into account of the remainder B through a self-energy.

24.4 Fano Line Shapes

This model gives a successful qualitative description of resonances like the state $2s2p\,^1P$ of He. Like all the states of He with both electrons excited, this one is unbound; it can be considered as a temporary bound state, with an energy of about 60 eV, well above the ionization threshold at about 25 eV above the ground state. It broadens by interacting with the autoionization continuum. Such situations are common in spectroscopies (e.g., electro energy loss, optical absorption). A common feature of many such spectra is a characteristic skew resonant line shape sitting on a non-resonant continuum. The Green's function G_{00} may be rewritten as

$$G_{00} = \frac{e^{i\Delta}}{\sqrt{[\omega - \epsilon_0 - \Sigma_1]^2 + \Sigma_2^2}},$$

where $\Sigma_1 = Re\,\Sigma(\omega)$, $\Sigma_2 = Im\,\Sigma(\omega)$ and

$$\tan(\Delta) = \frac{\Sigma_2}{\omega - \epsilon_0 - \Sigma_1}.$$

Thus, Δ is the phase of G_{00}. Since for large $|\omega|$ $G_{00} \sim \frac{1}{\omega}$, Δ changes by π across the resonance. We may rewrite Eqs. (24.23) and (24.12) in terms of Δ:

$$\begin{aligned} \langle 0 | \lambda_\omega \rangle &= \frac{\sin(\Delta)}{\pi \rho V_\omega} \\ \langle k | \lambda_\omega \rangle &= \frac{1}{\rho^c}\left[\frac{V(\epsilon_k)\sin(\Delta)}{\pi V(\epsilon_\lambda)(\epsilon_\lambda - \epsilon_k)} - \cos(\Delta)\delta(\epsilon_\lambda - \epsilon_k)\right]. \end{aligned} \tag{24.25}$$

Let $\hat{T}$ denote the operator of the transition from an initial state (for instance, the ground state of He) to the resonant H eigenstate $|\lambda_\omega\rangle$. One finds that

$$\langle i|\hat{T}|\lambda_\omega\rangle = -\cos(\Delta(\omega))\langle i|\hat{T}|k_\omega\rangle + \frac{\sin(\Delta(\omega))}{\pi\rho^c V_\omega}\langle i|\hat{T}|\tilde{0}\rangle, \tag{24.26}$$

where

$$|\tilde{0}\rangle = |0\rangle + P\sum_k \frac{V(\epsilon_k)}{\omega - \epsilon_k}|k\rangle$$

is the discrete state with a *halo* of continuum. As ω goes from below to above the resonance, $\tan(\Delta)$ diverges, $\sin(\Delta)$ is even and $\cos(\Delta)$ is odd; therefore, the line shapes can be asymmetric. For a simple qualitative analysis, Fano proposed considering

$$q = \frac{\langle i\hat{T}|\tilde{0}\rangle}{\pi\rho^c V_\omega\langle i|\hat{T}|k_\omega\rangle} \tag{24.27}$$

as a constant. The *reduced line shape* is defined as

$$f = |\frac{\langle i|\hat{T}|\lambda_\omega\rangle}{\langle i|\hat{T}|k_\omega\rangle}|^2. \tag{24.28}$$

It turns out that

$$f = \sin^2(\Delta)|-\cot(\Delta)+q|^2 = \frac{|-\cot(\Delta)+q|^2}{1+\cot^2(\Delta)}. \tag{24.29}$$

Away from the resonance, the denominator represents the background due to transitions to the unperturbed continuum. Rather than plotting versus ω, Fano proposed plotting versus the *reduced energy*

$$E = -\cot(\Delta) = \frac{\omega - \epsilon_0 - \Sigma_1}{\pi\rho^c V_\omega^2}; \tag{24.30}$$

in this way, one can compare the shapes of resonances centered at different energies with different widths. We arrive at the celebrated Fano line shape. For q = 0, it shows an *anti-resonance*, for q = 1, it becomes step-like, and for q = 3 it is a resonance with the typical asymmetry. Note that for E = −q, the transition is forbidden. For large q, the transition to the discrete state dominates and the shape tends to be symmetric. The Fano treatment is an example of the art of extracting the essential information from a phenomenological model that is by nature qualitative or semi-quantitative (Fig. 24.1).

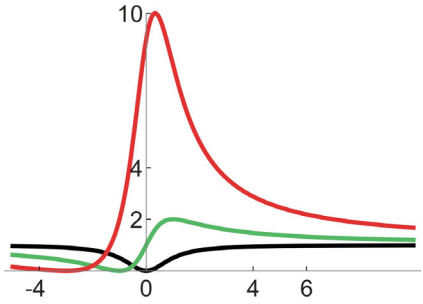


Fig. 24.1 The Fano line shapes. Black, q = 0; Green, q = 1; Red, q = 3

24.5 Anderson Model and Hubbard Model

The above discussion allows us to understand more about the Anderson impurity model which we have already talked about in Chap. 19.

$$H_A = \sum_{k,\sigma} \epsilon_k \hat{n}_{k\sigma} + \epsilon_0 \hat{n}_{0\sigma} + \sum_k [V_k c^\dagger_{k\sigma} c_{0\sigma} + h.c.] + U n_{0+} n_{0-}. \tag{24.31}$$

This idealized Hamiltonian represents a non-degenerate valence level of atom at site 0 interacting with a band continuum, including an interaction U between spin up and down states and was introduced to model ferromagnetic impurities in metals that in some conditions can develop a localized magnetic moment, affecting the conductivity of the system.[4] Under some conditions, this model is solved exactly,[5] but here we have space only for a simple argument. For U=0 this is a Fano model and is symmetric for spin up and down. Therefore, any magnetic moment is due to the U term. Then, to make progress, we may introduce a Hartree–Fock approximation in which the Hamiltonian for spin up electrons reads

$$H_{A,+} = \sum_{k,+} \epsilon_k \hat{n}_{k+} + \epsilon_0 \hat{n}_{0+} + \sum_k [V_k c^\dagger_{k+} c_{0\sigma} + h.c.] + U n_{0+} \langle n_{0-} \rangle, \tag{24.32}$$

(and of course for spin down electrons, + and − are interchanged). Now, without doing calculations, we can understand something qualitatively. It is clear that if $U = 0$ the filling of the site ϵ_0 is the same for both spins; hence, if the level is nearly empty or nearly full, and/or U is small enough, no localized moment arises, since there is a kinetic energy cost implied in filling spin up more than spin down or spin up less than spin down. So the self-consistent solution is symmetric and there is no localized moment. On the other hand, if U is large and the level filling is about half, there is a gain in energy in having spin up more filled than spin down (or the reverse) and

[4] The Kondo effect is an interesting phenomenon producing a resistance minimum as a function of temperature.

[5] P B Wiegman and A M Tsvelick, J. Phys. C: Solid State Phys. 16 (1983) 2281–2319

we can expect a pair of magnetic solutions. These considerations are confirmed by the explicit model calculations. About the same time (1963) John Hubbard , Martin Gutzwiller and Junjiro Kanamori independently proposed a model Hamiltonian of the form

$$H = -t \sum_{i,\sigma} (c^{\dagger}_{i,\sigma} c_{i+1,\sigma} + c^{\dagger}_{i+1,\sigma} c_{i,\sigma}) + U \sum_{i} \hat{n}_{i,+} \hat{n}_{i,-}, \tag{24.33}$$

representing a chain of sites with the same interaction U between opposite spins on each site; this model and many extensions to degenerate orbitals and higher dimensions with various geometries have been extensively studied. Applications of these models will be useful in Chap. 25.

References

Davis, L.C., Feldkamp, L.A.: Phys. Rev. B **15**, 2961 (1977)
Fano, U.: Phys. Rev. **124**, 1866–1878 (1961)

Chapter 25
Quantum Statistical Mechanics

The insightful classical approach by Gibbs (Sect. 5.5) in the quantum context fully reveals its power.

25.1 Temperature, Density Matrix and Decoherence Problems

Since thermal excitations are unavoidable, it is often impossible to prepare any sample in a *pure state*, that is, in the n-th eigenstate $\{|\psi_n\rangle\}$ of the system Hamiltonian $\hat{H}$. If we take into account the fact that at a finite temperature T the system can be thermally excited, the actual mean value of an operator $\hat{A}$ is

$$\langle \hat{A} \rangle = \sum_n P_n \langle \psi_n | \hat{A} | \psi_n \rangle, \tag{25.1}$$

where P_n is the probability of finding the system in $|\psi_n\rangle$. Quantum averages and thermal averages with P_n must be taken together.[1] To this end, one introduces the *density operator*

$$\hat{\rho} = \sum_n P_n |\psi_n\rangle \langle \psi_n|, \tag{25.2}$$

[1] In fact, both averages are performed by integrating along a generalized path, in the Keldysh formalism.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_25

and the mean (25.1) is given by

$$\langle \hat{A} \rangle = Tr(\hat{\rho}\hat{A}) \equiv \langle \psi_n | \hat{\rho}\hat{A} | \psi_n \rangle, \tag{25.3}$$

with $Tr(\hat{\rho}) = \sum_m \langle \psi_m | \hat{\rho} | \psi_m \rangle = 1$ and $\hat{\rho}^\dagger = \hat{\rho}$.

For a one-dimensional example, let $|k\rangle$ be the De Broglie wave with momentum $p = \hbar k$; then, we can make a coherent quantum superposition of $\langle x|k\rangle$ and $\langle x| - k\rangle$; the wave function might be $\psi(x) = \frac{1}{\sqrt{\pi}} \cos(kx)$. This state can be obtained by shooting particles with a gun against a potential wall at $x = 0$. The wave function vanishes by quantum interference when kx is an integer times π. This may be contrasted with the mixed state given by $\rho = \frac{1}{2}(|k\rangle\langle k| + |-k\rangle\langle -k|)$, which may be obtained by shooting particles with two guns (k from the left and -k from the right). A detector picks one or the other with the same probability, and no interference takes place. At finite temperatures, some loss of coherence may result from thermal fluctuations. But this is only the beginning of the story.

Actually Eq. (25.1) is contradictory, since it allows the system to have temperature T; however, it neglects the interaction of the system with the environment which is causing the system to thermalize. The (weak) excuse is that the environment is too large, and we are trying to study the system. So we say that there is *decoherence*. This really means that the real system is larger than the model we are considering, or, in other terms, the true Hamiltonian is much larger than our model. In more advanced texts, the reader can find the master equation and other approaches that have been proposed in the literature to include the system–environment interactions approximately and study the transition to classical behavior.[2]

25.2 Quantum Canonical Ensemble

In the canonical ensemble, at temperature T the probability that the system energy is one of the eigenvalues ϵ_n is $\frac{1}{Z} e^{-\frac{\epsilon_n}{KT}}$, where $\frac{1}{Z}$ is a normalization factor, in perfect analogy with the classical theory. While in classical Statistical Mechanics, we defined the canonical partition function (exrefeq5.42) as

$$Z = \int e^{-\beta H_S(p,q)} d\Gamma,$$

in terms of a phase space integral, the quantum *partition function* is

$$Z = Tr[e^{-\beta \hat{H}}] = \sum_{n=0}^{\infty} e^{-\frac{\epsilon_n}{KT}}. \tag{25.4}$$

[2] See, e.g., arXiv;2306.13717 [quant-ph] 23 June 2023.

So, at temperature T, $\frac{1}{Z}e^{-\frac{\epsilon_n}{KT}}$ is the probability of having a system at level n. According to (5.43), the internal energy U is given by

$$\overline{H} = -\frac{1}{Z}\frac{\partial Z}{\partial \beta} = -\frac{\partial \ln(Z)}{\partial \beta}. \tag{25.5}$$

Adding a constant to H_S modifies Z but leaves the physics unchanged. Z allows us to calculate averages. The mean energy is

$$\overline{E} = \frac{\sum_{n=0}^{\infty} \epsilon_n e^{-\frac{\epsilon_n}{KT}}}{\sum_{n=0}^{\infty} e^{-\frac{\epsilon_n}{KT}}} = \frac{Tr[He^{-\beta H}]}{Z} = -\frac{\partial}{\partial \beta} LogZ. \tag{25.6}$$

Let us work out this formula for the harmonic oscillator. Z_{osc} is the sum of a geometrical series: therefore,

$$Z_{osc} = \sum_{n=0}^{\infty} e^{-\frac{-n\omega\hbar}{KT}} = \left[1 - e^{-\beta\hbar\omega}\right]^{-1}. \tag{25.7}$$

25.3 Ideal Fermi Gas and Ideal Bose Gas

Suppose we have 2 particles and each of them has the same probability of being in one of two single-particle states A and B. What is the probability $P_{together}$ of having both in the same state? It depends on the statistics.

Classical Physics considers only *distinguishable particles*, since (at least, in principle) one can always mark them somehow and follow their trajectories with continuity. There are 4 cases:

$$\begin{cases} \psi_A(1)\psi_A(2) \\ \psi_A(1)\psi_B(2) \\ \psi_B(1)\psi_A(2) \\ \psi_B(1)\psi_B(2) \end{cases}$$

$\Longrightarrow P_{together} = \frac{1}{2}$. In Quantum Physics, identical particles are *indistinguishable*; there is no way to identify them and they cannot be recognized when observed again. This makes a lot of difference.

For *Fermions*, there is only one possibility:

$$\psi_A(1)\psi_B(2) - \psi_B(1)\psi_A(2)$$

$\Longrightarrow P_{together} = 0$. For *Bosons*, there are 3 cases:

$$\left\{ \begin{array}{r} \psi_A(1)\psi_A(2) \\ \psi_A(1)\psi_B(2) + \psi_B(1)\psi_A(2) \\ \psi_B(1)\psi_B(2) \end{array} \right.$$

$\Longrightarrow P_{together} = \frac{2}{3}$. In the case of photons, the enhanced probability of finding two of them in the same quantum state rather than in different states is called *photon bunching*. In other terms, there is quantum bunching when the second-order correlation function

$$\gamma(\tau) = \frac{\langle I_1(t) I_2(t+\tau) \rangle}{\langle I_1 \rangle \langle I_2 \rangle} \tag{25.8}$$

between the light intensities measured by two detectors is larger than 1 for small τ. This is not expected classically, but occurs in the lab when the light of a lamp is split into two by a half-silvered glass and the detector detects photons.[3]

Now suppose we have a gas of $N >> 1$ identical independent particles (perfect gas). Using the microcanonical ensemble, we suppose it is isolated with total energy E. The system is in a statistical (not quantum) superposition of quantum states, and those leading to the same macroscopic behavior have the same probability. The principle is the one formulated by Gibbs:

> *The basic assumption of Statistical Mechanics is* :
> *all the microstates that correspond to the same*
> *macroscopic parameters have the same probability*
> *in equilibrium.*

One can measure the energies ϵ of individual molecules in the gas. They turn out to be distributed according to a law $n(\epsilon)$.

To define a microstate, we work in the μ space, which is the phase space of the single particle, with coordinates x, y, z, p_x, p_y, p_z. We make this space discrete, that is, we divide it into small cells. More precisely, the size of a cell must be macroscopically negligible, yet it must be large enough to accommodate the coordinates of many molecules at any given instant of time, and the ith cell must contain $g_i \gg 1$ degenerate quantum states for a single particle. We also make the single-particle energies discrete by dividing the ϵ axis into small intervals. Next, we consider all the possible ways to distribute the N gas molecules among the cells with n_i molecules in the ith cell, with the constraint that the number of molecules and the total energies are given, that is,

$$\sum n_i = N, \tag{25.9}$$

[3] Photon antibunching has also been reported in special circumstances. In 1977 Mandel, Kimble and Dagenais produced antibunching using as source a single atom. The photons were emitted by a transition that could not be repeated before taking the atom back to the excited state.

$$\sum \epsilon_i n_i = E. \tag{25.10}$$

Since the particles are indistinguishable, the exchange of particles between different states makes no sense. We first seek to find how many microstates realize a given distribution $n_1, n_2, n_3 \ldots$ and then we obtain the most probable distribution. For classical Boltzmann particles, one obtains

$$\overline{n}_r = g_r e^{-\beta(\epsilon_r - \mu)},$$

where μ is the chemical potential. We get different solutions for Bose and Fermi cases.

25.3.1 Bose–Einstein Statistics

The Boltzmann distribution (5.77) was obtained by looking for the most probable distribution of *distinguishable* particles in phase space (Chap. exrefchap5). The fact that in principle the quantum particles cannot be labeled is not a subtlety but makes a lot of difference except at high energies, as we shall see.

The Bose-Einstein Statistics assigns the equilibrium value $\overline{n}_i$ of the number of Bosons in cell i. A microstate is specified when all the mean numbers of bosons in each quantum state are assigned. The equilibrium condition corresponds to the maximum number W of microstates that can be attained by varying the population n_r of cell r compatibly with the constraints. Besides the total energy E, the number N of Bosons is fixed in the case of atoms or molecules, but not in the case of Photons, that can be absorbed or emitted. Therefore

$$\frac{\partial}{\partial n_r} [\ln W - \alpha N - \beta E] = 0, \tag{25.11}$$

where α and β are Lagrange multipliers, but α=0 in the case of photons. Let us develop in the case of atoms. The probability of a distribution of the energies such that n_r atoms are in the cell r depends on the number g_r of degenerate states in that cell. So we set

$$W = \prod_r W_r, \tag{25.12}$$

where W_r is the number of ways one can distribute n_r particles in cell r among the degenerate g_r states. This is the same problem as finding the number of ways to put n_r marbles in g_r drawers. For 2 marbles in 3 drawers, one finds 6 ways:

$$|| \cdot \cdot \qquad | \cdot | \cdot \qquad | \cdot \cdot |$$

$$\cdot || \cdot \qquad \cdot | \cdot | \qquad \cdot \cdot ||$$

The first diagram puts 2 marbles in drawer 3, the second diagram puts one in the second and one in the third, and so on. The positions of the vertical lines | correspond to the pairs (1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4). In general, one obtains a configuration for each way of aligning the dots with the $g_r - 1$ vertical lines, so

$$W_r = \binom{n_r + g_r - 1}{g_r - 1} \approx \binom{n_r + g_r}{g_r}, \tag{25.13}$$

since $g_i \gg 1$. Then,

$$\frac{\partial}{\partial n_i} \ln W = \frac{\partial}{\partial n_i} \sum [(n_r + g_r) \ln(n_r + g_r) - n_r \ln n_r - g_r \ln g_r] = \ln \frac{n_i + g_i}{n_i}.$$

So, the condition (25.3.1) is satisfied by

$$\ln \frac{n_i + g_i}{n_i} - \alpha - \beta \epsilon_i = 0;$$

this yields the Bose–Einstein (BE) distribution (number of Bosons in the i-th cell of phase space)

$$\boxed{\overline{n}_i = \frac{g_i}{e^{\alpha + \beta \epsilon_i} - 1} \equiv \frac{g_i}{e^{\beta(\epsilon_i - \mu)} - 1}.} \tag{25.14}$$

The parameter α for atoms is fixed by $\sum n_i = N$, and one sets $\alpha = \beta\mu$ where μ is the chemical potential. There is no risk of a vanishing denominator, and $e^{\beta(\epsilon_i - \mu)} > 1$, which implies that the occupied levels are above the chemical potential μ by a finite gap.

As the notation suggests, $\beta = \frac{1}{K_B T}$; indeed, for $\beta \epsilon_i \gg 1$, we recover the Boltzmann distribution.

Ideal Gas of Bosonic Atoms

Above a critical temperature T_C, the particles occupy all the cells in such a way that every single state gives a negligible contribution, in analogy with the Boltzmann distribution; however, T_C is the onset of the Bose–Einstein condensation. This marks a dramatic departure from the classical behavior. Below T_C, a number n_0 of particles, which is a finite fraction of N, is in the state with $\epsilon = 0$. This is the condensate. As a consequence, when we work out the sum over states as an integral, we must exclude the population of $\epsilon = 0$ from the integral.

Let us consider $T > T_C$ first. In order to obtain the number of atoms in the cell in phase space, whose volume is h^3, one must multiply the Bose distribution by $\frac{(2s+1) d^3 p dV}{h^3}$, and the condition (25.9) is readily found by integrating over the angles:

$$N = \frac{4\pi V (2s + 1) \sqrt{2m^3}}{h^3} \int_0^\infty \frac{\sqrt{\epsilon} d\epsilon}{e^{\beta(\epsilon - \mu)} - 1}. \tag{25.15}$$

The chemical potential

Let us find μ. To this end, we introduce De Broglie's thermal wavelength, which is defined as

$$\lambda(T) = \frac{h}{\sqrt{2\pi m K_B T}}, \tag{25.16}$$

then define the fugacity $z = e^{\beta\mu}$. and define the integral

$$g_{\frac{3}{2}}(z) = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{\sqrt{x}dx}{z^{-1}e^x - 1}. \tag{25.17}$$

Now we can rewrite the condition (25.15) in the form

$$\frac{N}{V(2s+1)} = \frac{g_{\frac{3}{2}}(z)}{\lambda^3(T)}. \tag{25.18}$$

This is the equation that must be solved for the fugacity z and so yields μ.

Below the critical temperature

When T decreases toward T_C, we reach the critical point where the chemical potential $\mu = 0, z = 1$ and $g_{\frac{3}{2}}(1) = \zeta(\frac{3}{2}) \approx 2.612$, where ζ is the Riemann zeta function. Hence, the critical temperature is given by

$$T_C = \left(\frac{N}{(2s+1)V\zeta(\frac{3}{2})}\right)^{\frac{3}{2}} \frac{h^2}{2\pi m K_B}. \tag{25.19}$$

For $T < T_C$, Eq. (25.15) gives

$$N - n_0 = \frac{4\pi V(2s+1)\sqrt{2m^3}}{h^3} \int_0^\infty \frac{\sqrt{\epsilon}d\epsilon}{e^{\beta(\epsilon-\mu)} - 1}, \tag{25.20}$$

where $n_0 = \frac{1}{e^{-\beta\mu}-1} = \frac{z}{z-1}$.

Liquid He 4

Liquid He 4 exists in two distinct phases. Above a critical temperature ($T_C = 2.19\,^\circ$K at low pressures, but $2\,^\circ$K at 10 atmospheres), He I behaves as an ordinary liquid. Below T_C, He II is still a liquid, but a very unusual one. Pyotr Kapitsa, John F. Alen and Don Misener discovered, in 1937, that He 4 becomes a superfluid, with zero viscosity and zero entropy. This is a manifestation of the Boson condensation. It does not become a solid at lower temperatures, because the zero point oscillations keep the would-be solid consistently molten unless under a pressure of about 25 atmospheres. Moreover, He II is a superfluid, with quantized vortices. It has a very high thermal conductivity and heat propagates as *second sound* waves. He II has a complex spectrum of elementary excitations. The properties of the ideal Bose gas are

somewhat spoiled by the interactions, but a partial condensation occurs and a fraction of the atoms go to a single state, the one of lowest energy *(Bose condensation)*.

Ultracold Atoms

The principle of laser cooling is based on the Doppler effect. Atoms absorbing and emitting photons recoil in a way that depends on their velocities, and the net effect tends to slow them down. A difficulty is that when the atom becomes slower, it tends to be out of resonance with the laser, but a technique based on an inhomogeneous magnetic field was devised to keep the atoms in resonance. The Bose–Einstein condensation of supercool gases, based on this device, was obtained experimentally in the '90s. William Phillips, Claude Cohen-Tannoudji and Steven Chu received the 1997 Nobel prize for this achievement and for opening a new chapter of Physics at ultra-low temperatures. Magneto-optical traps using laser cooling produce samples of cold atoms in the micro-Kelvin range. This technique lends itself to quantum simulations, thanks to the excellent experimental control of the samples.

25.4 Black Body and Photons

In the case of Photons, the condition $\sum n_i = N$ is missing (the number of photons is variable), and the Lagrange multiplier is $\alpha = 0$. So, the mean number of photons in each normal mode (energy $\epsilon_\omega = \hbar\omega$) is

$$\overline{n}_\omega = \frac{1}{e^{\beta\epsilon_\omega} - 1},$$

and each normal mode contains the energy

$$\overline{E}_\omega = \frac{\epsilon_\omega}{e^{\beta\epsilon_\omega} - 1}. \tag{25.21}$$

The $\boldsymbol{k}$ points form a lattice in reciprocal space, and each corresponds to two normal modes (two polarizations). In the continuum limit, the number of modes in a volume of $\boldsymbol{k}$ space is obtained dividing by the volume associated with each mode:

$$\sum_k \to \frac{L^3}{(2\pi)^3} \int_k d^3k. \tag{25.22}$$

Let $V = L^3$ denote the volume of the photon gas; the number of modes in dk is $2\frac{V}{8\pi^3} 4\pi k^2 dk$. Since $\omega = ck$, the number of modes in $d\omega$ is $\frac{V\omega^2 d\omega}{\pi^2 c^3}$. Then, the energy at ω is

$$d\overline{E}_\omega = \hbar\omega \frac{1}{e^{\beta\epsilon_\omega} - 1} \frac{V\omega^2 d\omega}{\pi^2 c^3} \equiv V u(\omega) d\omega. \tag{25.23}$$

Hence, we get Planck's law for the energy density:

$$u(\omega) = \hbar\omega \frac{1}{e^{\beta\epsilon_\omega} - 1} \frac{\omega^2}{\pi^2 c^3},$$

or

$$\boxed{u(\nu) = \frac{8\pi^2 h\nu^3}{c^3} \frac{1}{e^{\frac{h\nu}{K_B T}} - 1}} \tag{25.24}$$

which historically was the starting point (for Planck) of Quantum Theory in the year 1900. A series expansion gives $u(\nu) = \frac{8\pi\nu^2 K_B T}{c^3} - \frac{4\pi\, \mathrm{h}\nu^3}{c^3} + \frac{2\pi h^2\nu^4}{3c^3 K_B T} + O(\nu^5)$ and the first term is the classical Rayleigh–Jean formula (Chap. exrefchap5) which is perfect at low frequencies but leads to the *ultraviolet catastrophe*. The physics is clarified: since the field is quantized, the thermal energy $K_B T$ can hardly excite the photon modes with $h\nu > K_B T$, and these modes remain frozen.

The simple Planck formula, which holds true for all the known electromagnetic spectra, from radio waves to γ rays, that is, from nearly zero to more than 10^{19} Hertz, has the remarkable property that when the black body is in motion relative to the observer, the spectrum is Doppler-shifted, but still appears as a black body spectrum, albeit, at a different temperature (hotter if the body is approaching, cooler if it is receding). The basic reason is simple: if it is black, it remains black.

The maximum of the distribution ω_{max} is proportional to the temperature. This is Wien's law. Integrating over frequency, we obtain the Stefan–Boltzmann: law:

$$U = \int u d\omega = \frac{\hbar}{\pi^2 c^3} \int_0^\infty \frac{\omega^3 d\omega}{\exp[\hbar\frac{\omega}{K_B T}] - 1}.$$

Setting $x = \beta\hbar\omega$,

$$U = \frac{\hbar}{\pi^2 c^3} \left(\frac{K_B T}{\hbar}\right)^4 \int_0^\infty \frac{x^3 dx}{e^x - 1}. \tag{25.25}$$

Now,

$$\int_0^\infty \frac{x^3 dx}{e^x - 1} = \frac{\pi^4}{15}$$

and $U \sim T^4$, but the constant of proportionality, which could not be determined by thermodynamics, is fixed by the Quantum theory.

25.4.1 Einstein A and B Coefficients and the Principle of the LASER

Already in 1917, A. Einstein used Planck's law to establish relations between absorption and emission of radiation by a molecule at temperature T. The rate of stimulated absorption is $W_a = B_a u(\omega)$, where B_a is the absorption coefficient and u is given by 25.24. The rate of stimulated emission is $W_{se} = B_{se} u(\omega)$, where B_{se} is the stimulated emission coefficient. The concept of a stimulated emission was introduced for the first time. Letting A denote the spontaneous emission rate, the total emission rate is $W_{tot} = A + B_{se} u(\omega)$. Introducing the numbers of molecules N_u and N_l in the upper and lower levels, respectively, the Boltzmann distribution requires that $\frac{N_u}{N_l} = \exp(-\frac{h\nu}{K_B T})$. In equilibrium, $N_l B_{sa} u(\omega) = N_u (A + B_{se} u(\omega))$; this gives us

$$u(\omega) = \frac{\frac{A}{B_{sa}}}{\frac{N_l}{N_u} - \frac{B_{se}}{B_a}}.$$

Comparing with Planck's law, one finds that $B_a = B_{se}$ and $A = \frac{8\pi h\nu^3}{c^3} B$. Normally, a light wave is absorbed by matter, but if an excited state is more populated than the ground state, *the stimulated emission prevails on absorption and the light is amplified*. The population inversion can be generated by optical or electric excitation. The Light Amplification Stimulated Emission of Radiation (LASER) is based on this principle, but in order to achieve a stable population inversion, important practical problems had to be solved. Fundamental work was done by Charles H. Townes[4] and Arthur Leinard Schawlow. The first LASER built by Theodore Maiman at Huges Research Laboratories only started to operate in 1960. The stimulated emission is produced in a cavity and the frequency is one of its resonant frequencies; in this way all the photons produced by the laser correspond to a single mode and are in phase, with the result that the laser beam has a high degree of coherence and directionality. Initially it was regarded as a useless curiosity, just an interesting toy for scientists; now, the applications in surgery and everyday life are countless and need not be explained here.

Moreover, in 1971 at Stanford University, John Madey built and operated the first free electron laser (FEL) based on a different scheme, which uses special Relativity rather than quantum mechanics. The photons are emitted by 43 MeV electrons (from an accelerator) crossing a wiggler which is a periodic arrangement of magnets with alternating poles. The wiggler is inside an optical cavity. Then, standing waves are formed and the whole system emits X-ray coherent radiation. This is used for several applications including basic research, surgery and (of course) military (star wars).

[4] Charles Hard Townes (July 28, 1915–January 27, 2015) was an American physicist. He worked on the theory and application of the maser. He shared the 1964 Nobel Prize in Physics with Nikolay Basov and Alexander Prokhorov.

25.4.2 *Specific Heats of Solids*

The classical equipartition theorem assigns to every oscillator in 3 dimensions the energy $E_{osc} = 3K_B T$. Therefore, the contribution of each oscillator to the specific heat $C_V = \frac{\partial E_{osc}}{\partial T}$ should be $3K_B$. This is *grosso modo* true around room temperature (the Dulong and Petit law), but specific heats drop at low temperatures.

By 1907, well before the development of Quantum Theory, Albert Einstein had already proposed a quantum model of the specific heats of solids, in which the quantized atomic vibrations (phonons) give to the energy a contribution

$$E_{phon} = \frac{n\hbar\omega}{e^{\beta\hbar\omega} - 1},$$

where n is the number of oscillators and ω their pulsation; the phonons are Bosons. The specific heat is given by

$$C_V \equiv \frac{\partial E_{phon}}{\partial t} = \frac{n(\hbar\omega\beta)^2 K_B e^{\beta\hbar\omega}}{(e^{\beta\hbar\omega} - 1)^2}.$$

For $\beta \to 0$ (high temperatures) this agrees with the classical Dulong and Petit law. Below the Einstein temperature $T_E = \frac{\hbar\omega}{K_B}$, the drop $C_V \to 0$ for $T \to 0$ is in qualitative agreement with experiment.

25.5 Fermi–Dirac Statistics

Like in the classical case (5.77) and the Bose-Einstein case that we have just seen, the statistics suitable for Fermions must be obtained by looking for the most probable distribution of particles in phase space. We must treat the fermions as *indistinguishable* and the occupation number of each quantum state can only be 0 or 1. We divide the phase space into cells; each cell must be macroscopically small, but cell i will contain $g_i \gg 1$ states. Cell i contains n_i particles in g_i states, and there are

$$W_i = \binom{g_i}{n_i}$$

ways to put n_i particles therein. Since the exchange of identical Fermions has no meaning, a microstate is assigned once the numbers $n_1, n_2, n_3, \ldots$ of particles in the cells of energy $\epsilon_1, \epsilon_2, \epsilon_3 \cdots$ is given.

To find the most probable distribution, the number W of ways, we can realize the distribution must be maximized under the constraints $\sum_i n_i = N$, $\sum_i \epsilon_i n_i = E$. Since the cells are statistically independent,

$$W = \prod_i W_i = \prod_i \frac{g_i!}{n_i!(g_i - n_i)!}$$

and Stirling's formula $\ln n! \approx n \ln n$ gives us

$$\ln W \approx \sum_i \{g_i \ln g_i - n_i \ln n_i - (g_i - n_i) \ln(g_i - n_i)\}. \tag{25.26}$$

We need the maximum for fixed E and N. By the method of Lagrange's multipliers, we impose the vanishing of

$$\frac{\partial}{\partial n_r}\left[\sum_i \{g_i \ln g_i - n_i \ln n_i - (g_i - n_i) \ln(g_i - n_i) - \alpha \sum n_i - \beta \sum \epsilon_i n_i\right].$$

Since 1 is negligible compared to n_i, one finds

$$-\ln n_r - (-) \ln(g_r - n_r) - \alpha - \beta\epsilon_r = 0.$$

Hence, $\ln\left[\frac{g_r - \overline{n}_r}{n_r}\right] = \alpha + \beta\epsilon_r$ and this implies the Fermi–Dirac distribution

$$\boxed{\overline{n}_r = \frac{g_r}{e^{\alpha+\beta\epsilon_r} + 1} \equiv \frac{g_r}{e^{\beta(\epsilon_r - \mu)} + 1}}, \tag{25.27}$$

where μ is the chemical potential. For large $\epsilon_r - \mu$, this reduces to the Boltzmann distribution with $\beta = \frac{1}{K_B T}$. Only a sign in the denominator distinguishes the Fermi from the Bose distribution.

At absolute zero, the Fermi function $f(x) = \frac{1}{e^{\beta x}+1}$ becomes a step function, and $\mu = E_F$, since the step is at the *Fermi level*, or highest occupied level. The step becomes gradual at finite temperatures, but for a typical metal $E_F =$ some eV, and many thousand degrees are needed to round the step significantly (room temperature corresponds to 25 meV).

25.5.1 Fermi Gas

The Fermi distribution describes the occupation of electronic states in solids. The *Fermi gas* is the perfect gas of Fermions with a given uniform density ρ. If interactions are considered, it becomes a *Fermi liquid*, characterized by plasma oscillations (e.g., in metals) and in He 3 by the propagation of hydrodynamic waves called zero sound, having frequency of the order of $q v_F$, where q is the wave vector and v_F is the Fermi momentum divided by the electron mass.

1 dimension ($d = 1$)

Suppose there are $N \gg 1$ electrons in a *box* of length L with periodic boundary conditions[5] ($x = L$ is identified with $x = 0$; if you like, you can think of a ring). The density is $\rho = \frac{N}{L}$ and the wave functions have the space-dependent factor

$$\psi_k(x) = \frac{1}{\sqrt{L}} e^{ikx};$$

the boundary conditions $e^{ikL} = 1$ allow for the wave vectors

$$k_n = \frac{2\pi n}{L}. \tag{25.28}$$

The interval between two allowed wave vectors is

$$\Delta k = \frac{2\pi}{L}.$$

The N electrons in the ground state of the Fermi gas fill the levels according to the *aufbau* principle, up to the *Fermi level*; it corresponds to the wave vector $k = k_F$, to the momentum $p_F = \hbar k_F$ and to the energy $\epsilon_k = \frac{p_F^2}{2m}$. Between $k = 0$ and $k = k_F$ there are N spin-orbitals, that is, $N/2$ orbitals:

$$N = 2 \sum_{k<k_F} 1.$$

To compute k_F, one goes over to a continuous description and

$$N/2 = \sum_{k \le k_F} \to \int_0^{k_F} \frac{dk}{\Delta k} = \frac{k_F}{\Delta k}, \tag{25.29}$$

therefore, k_F is given by

$$k_F = \Delta k \frac{N}{2} = \frac{\pi N}{L} = \pi \rho, \tag{25.30}$$

$d = 3$

In 3 dimensions, this becomes the old *Sommerfeld theory* of metals: let us consider a gas with density $\rho = \frac{N}{L^3} \equiv \frac{N}{V}$. Instead of (25.29), in $3d$, one writes

$$\sum_{k<k_F} = \frac{L^3}{(2\pi)^3} \int_{k<K_F} d^3k. \tag{25.31}$$

[5] For large L, the properties become independent of the size of the box.

Again, we put 2 electrons in every orbital in order of increasing energy until all the electrons have been used (the *aufbau* method); in spherical coordinates,

$$N = 2\sum_{k<k_F} = 2\frac{V}{(2\pi)^3}\int_{k<k_F} d^3k = \frac{2\,V4\pi}{8\pi^3}\int_0^{k_F} k^2 dk,$$

and so

$$N = \frac{Vk_F^3}{3\pi^2}, \tag{25.32}$$

that is,

$$k_F = (3\pi^2\rho)^{\frac{1}{3}} \sim 3.093\ \rho^{\frac{1}{3}}.$$

Fermi Gas Properties of Simple Metals

Elements with s or p valence electrons like Na, K and Al are called simple because qualitatively several properties can be described as by the Sommerfeld model, or by Jellium model, that I illustrate briefly in this paragraph. The Fermi gas with density ρ has a characteristic length, and the mean radius per particle r_0 is defined by

$$\frac{1}{\rho} = \frac{4\pi}{3}r_0^3. \tag{25.33}$$

Therefore, $r_0 \sim \frac{1}{k_F}$. Usually it is measured in Bohr radii introducing the Wigner–Seitz radius $r_s = \frac{r_0}{a_B}$. The mean distance between an electron and its nearest neighbor is $\sim r_0$. Cs has $r_s \sim 6$, while for Al, $r_s \sim 2$; the other metallic elements are in this range. The Fermi energy is

$$E_F = \frac{\hbar^2}{2m}(3\pi^2\rho)^{\frac{2}{3}}.$$

For $\rho = 0.1$ a.u. the Fermi energy is[6] $E_F \sim 1.03 a.u. \sim 28$ eV; the Fermi energies of metallic elements are typically 5–10 eV. The total energy of the gas is

$$E_{tot} = 2\sum_{k<K_F} \hbar^2\frac{k^2}{2m} = \hbar^2\frac{V}{2\pi^2 m}\int_0^{K_F} k^4 dk = \hbar^2\frac{V}{10\pi^2 m}{k_F}^5, \tag{25.34}$$

and the kinetic energy density

$$\frac{E_{tot}}{V} \sim \rho^{\frac{5}{3}} \sim r_0^{-5} \tag{25.35}$$

[6] In Atomic Units (a.u.) we set $\hbar = 1$, lengths are in Bohr radii (0.529 ρÅ) and energies in Hartrees (27.2 eV).

is a steep function of the density. Using (25.32), the mean kinetic energy per electron is found to be

$$\bar{\epsilon} = \frac{E_{tot}}{N} = \frac{3}{5} E_F.$$

For $\rho = 0.1$a.u., the energy density is about 0.06 a.u.

The Sommerfeld theory explains why the electronic contribution to the specific heat of metals grows linearly with the temperature T.

However, the energy density of the Fermi gas is positive (there is only kinetic energy) and the electrons must be confined in the metal by some box, otherwise they would escape and the metal should be unstable. While it is clear that electrostatic forces must be responsible somehow, the cohesion of metals is harder to understand than that of ionic crystals, which is readily explained classically. The Sommerfeld theory is not enough.

To describe metals, the *Jellium* model is widely used; in it, the electrons move over a uniform positive charge background. The negative charge of the electrons is exactly compensated by the background, and there is no direct electrostatic contribution to the ground state energy. However, there is also an exchange contribution. One can show that this can explain the cohesion of the metal in a limited range of ρ, which is close to the observed one. Within the Jellium model, one forgets many complications of real solids and can study more easily the phenomenon of screening, which is the attenuation of the potential of an external charge due to the presence of the medium. The bare potential of the external charge produces a redistribution of the electrons with the result that the repulsive effective potential is strongly reduced, except at small distances.[7] The many-body theory needed to study such phenomena is outside our scope. However, the end result is that one can introduce a wave vector and frequency-dependent dielectric function $\epsilon(q, \omega)$ such as the screened interaction is obtained by the Fourier-transformed static Coulomb interaction $V_C(\overrightarrow{q}, \omega) = \frac{4\pi e^2}{q^2}$ by a simple division:

$$V_s(\overrightarrow{q}, \omega) = \frac{4\pi e^2}{\epsilon(\overrightarrow{q}, \omega) q^2}. \tag{25.36}$$

Moreover, one finds that the mean energy per electron is

$$\bar{\epsilon} \sim \frac{2.21}{r_s^2} - \frac{0.916}{r_s} \tag{25.37}$$

where the energy is in 1 Rydberg = 13.59 eV. The first term is kinetic energy, while the second is exchange. Extra corrections, which become unimportant at small r_s, arise from correlation (that is, from the fact that electrons can reduce the mean repulsion by adopting a many-body wave function that is a mixture of many Slater determinants).

[7] I have no space here to discuss the Friedel oscillations,which anyhow have no practical effect.

The above oversimplified picture should not conceal the fact that, in solids, widely different phenomena can be caused by tiny energy differences. Magnetism is a very large and growing chapter of solid state physics. A weak effective attraction between the Fermions near the Fermi level can be the result of effective screening of the repulsion and small lattice and/or magnetic effects. The Fermions tend to form pairs called Cooper pairs at low temperature. This leads to superconductivity in which a macroscopic quantum wave function is formed (Sect. 26.5.7). The Cooper pairs are not Bosons but behave somewhat like Bosons, except that they do not produce a Bose condensate but a more subtle many-body ground state with perfect diamagnetism and zero resistance.

Unlike the classical gas, the Fermi gas tends to be perfect at high density, because the kinetic energy dominates. Everything becomes a metal under high enough pressure, and the interior of the planet Jupiter is believed to contain a deep ocean of metallic Hydrogen.

25.5.2 The Meitner–Auger Effect

This effect has been discovered experimentally in atoms in 1923 by the outstanding Austrian physicist Lise Meitner, often underestimated for a silly prejudice against women, and the mechanism clarified in 1925 by the great French physicist Pierre Auger. The Fermi distribution is realized in the ground state of the electron gas (i.e., non-interacting Fermi liquid, but is also a fair approximation for the conduction bands of some real metals (called simple metals for this reason) like Alkali metals and Al. Any state with a hole in a single-electron level below the Fermi level is an intrinsically unstable excited state and must decay somehow. Indeed, it can decay by filling the hole with a less bound electron and emitting a photon; this is the photoemission process. But it can also decay by a mechanism driven by the Coulomb interaction itself through a self-ionizing process, and this is the Auger process. It works as follows.

Let $|\Phi_i\rangle = c_h|\Phi_{gs}\rangle$ denote the state obtained from the N-electron ground state $|\Phi_{gs}\rangle$ by removing an electron in a level below the Fermi energy, often in a strongly bound core level; this excited state decays by filling the hole with a less bound electron and shooting a second electron out of the system. This is the Auger electron. As a consequence, the system is left with a pair of holes, both nearer in energy to the Fermi energy than the original hole. The whole process conserves the total energy.[8]

The final state will be of the form $|\Phi_f\rangle = c_k^\dagger c_{h1} c_{h2}|\Phi_{gs}\rangle$, where $c^\dagger(k)$ creates the free Auger electron, which is not interacting with the rest of the system any more, and $h1, h2$ are hole states. The transition is allowed by energy conservation if $|\Phi_f\rangle$ is degenerate with $|\Phi_i\rangle$ and the Coulomb interaction has a matrix element

[8] The final state can decay later again by the Meitner–Auger Effect, since no hole state is stable; however, we shall continue to treat it as final, which is adequate for what follows.

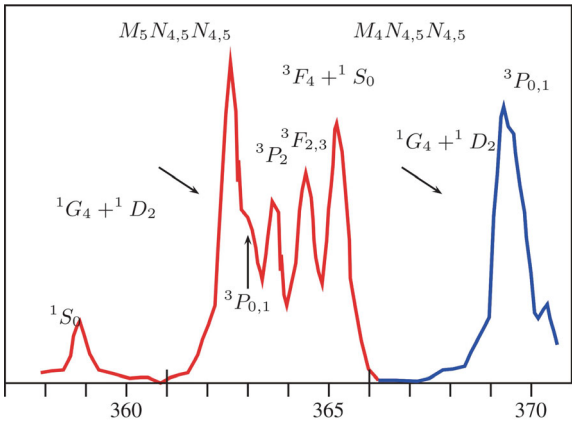


Fig. 25.1 Sketch of the Auger $M_5N_{4,5}N_{4,5}$ (red) and $M_4N_{4,5}N_{4,5}$ (blue) spectrum of Cd vapor, in arbitrary units, from measurements by H. Aksela and S. Aksela. Many of the multiplet terms are well-resolved and the intensities also convey useful information on the hole wave functions. The labels denote the final-state two-hole states, in the notation explained in Table 25.1. The assignments were done by intermediate coupling calculations of line positions and intensities. The intermediate coupling calculations include effects of the Coulomb interaction and of the relativistic spin–orbit coupling

Table 25.1 The atomic subshells are denoted by principal quantum number, the symbol s, p, d, ⋯ of the orbital angular momentum and a suffix showing the total one-electron angular momentum for one electron, as shown in the upper row. The corresponding spectroscopic notations for the same subshells are reported in the lower row

$1s_{\frac{1}{2}}$	$2s_{\frac{1}{2}}$	$2p_{\frac{1}{2}}$	$2p_{\frac{3}{2}}$	$3s_{\frac{1}{2}}$	$3p_{\frac{1}{2}}$	$3p_{\frac{3}{2}}$	$3d_{\frac{3}{2}}$	$3d_{\frac{5}{2}}$	$4s_{\frac{1}{2}}$	$4p_{\frac{1}{2}}$	$4p_{\frac{3}{2}}$	$4d_{\frac{3}{2}}$	$4d_{\frac{5}{2}}$	$4f_{\frac{5}{2}}$	$4f_{\frac{7}{2}}$
K	L_1	L_2	L_3	M_1	M_2	M_3	M_4	M_5	N_1	N_2	N_3	N_4	N_5	N_6	N_7

connecting $|\Phi_i\rangle$ with $|\Phi_f\rangle$. The German Gregor Wentzel already in 1927 obtained the transition rate

$$P_{if} = \frac{2\pi}{\hbar}|\langle\Phi|H_C|\Phi_f\rangle|^2 \tag{25.38}$$

through the Fermi Golden rule.

There may be many alternative hole states available to a system for decay, and the final-state holes can decay themselves by the same mechanism until the Fermi distribution is established. A plot of the Auger electron current versus kinetic energy is called the Auger spectrum; indeed, Auger spectroscopy is rich in information on the chemical state of the atom and on the local physics of the system. The Auger effect is in competition with radiative decay, which is generally faster in heavy atoms (X-ray emission). In modern theories, the production and the various decay channels of a core-hole in atoms is described as a coherent quantum process. In Auger spectra, one can observe the atomic multiplets of the two final-state holes, as shown in Fig. 25.1.

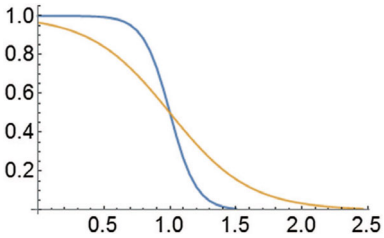


Fig. 25.2 Thick curve: Fermi function for $K_B T = 0.1\mu$; thin curve: Fermi function for $K_B T = 0.3\mu$

The red spectrum originates from the decay of a M_4 hole and the blue from a M_5 hole. The multiplet symbols that label each peak refer to the two-hole final state of the Auger transition. For example, ${}^3P_2$ means that the two holes are in a triplet state, with orbital angular momentum $\mathrm{L} = 1$ and total angular momentum $J = 2$ (Fig. 25.2).

25.5.3 Auger Line Shapes in Solids

Extensions of the Anderson model involving degenerate continua and degenerate localized state have been much studied. In particular, the degenerate Anderson model has been used to model the Auger Core-Valence-Valence spectra of solids. In the Auger process, a hole in a core level decays leaving the system with two holes in the valence band plus an electron—the Auger electron—which is detected and whose energy is analyzed. The energy distribution of the Auger electrons gives information on the final-state holes in the solid and their interaction, which is due to the screened interaction in the valence band. It was discovered experimentally by Powell that some Auger spectra were broad and similar to a self-convolution of the valence band density of states, while others seemed quasi-atomic.

It turns out[9] that an Anderson model and a Hubbard model represents well the situation. This is due to a quantum mechanical effect. If the repulsion is so strong that the two holes are outside the two-hole continuum, they become bound by the repulsion. This fact had been already noted by Kanamori[10] as a property of the Hubbard model, but Auger spectroscopy had changed an abstract result in a rather spectacular fact. In dealing with the Auger spectra, one must include relativistic effect and several complications of atomic physics. The Hubbard model with many modifications has been extensively studied in relation of magnetism and superconductivity, and there is a vast and interesting literature.

[9] Cini (1976) and Cini (1977); a similar theory was proposed by Sawatzky (1977) using the Hubbard model.

[10] Kanamori (1963).

25.5.4 Degeneracy Pressure

The Fermi distribution is a sharp step at absolute zero; with increasing $K_B T$, the step becomes a gradual drop. The chemical potential μ corresponds to a $\frac{1}{2}$ occupancy. At $K_B T > \mu$, many electrons are in the exponential tail of the distribution, and the Fermi gas resembles the classical one. In metals, under normal conditions, the step is sharp and the gas is *degenerate*. Its equation of state is quite unlike a classical gas. For fixed N, the energy (25.34) depends on volume:

$$E_{tot} = \hbar^2 \frac{V}{10\pi^2 m} (3\pi^2 N)^{\frac{5}{3}} V^{\frac{-5}{3}} = \eta \, V^{-\frac{2}{3}}, \tag{25.39}$$

where

$$\eta = \frac{\hbar^2}{10\pi^2 m} (3\pi^2 N)^{\frac{5}{3}}. \tag{25.40}$$

So, even at absolute zero, the Fermi gas has pressure, which is of quantum nature. One can evaluate this pressure through a thought experiment. When the volume expands adiabatically[11] by dV, the gas does a work $\delta L = P dV$; its energy changes by

$$dE_{tot} = -P dV. \tag{25.41}$$

From (25.39),

$$\frac{dE_{tot}}{dV} = -\frac{2}{3}\frac{E}{V}.$$

So, the *degeneracy pressure* is

$$P = \frac{2}{3}\frac{E_{tot}}{V} = \frac{2}{3}\hbar^2 \frac{(3\pi^2\rho)^{\frac{5}{3}}}{10\pi^2 m}. \tag{25.42}$$

The equation of state of the classical perfect gas reads as $P = N K_B T \rho$; the pressure vanishes at absolute zero and grows linearly with the density. Equation (25.42) predicts a very different trend, with a steeper increase of pressure, even at the lowest temperatures. This explains why metals are highly incompressible.

Degenerate Stars

The same law explains the stability of the degenerate stars (white dwarfs and neutron stars). White dwarfs are stellar remnants with a mass on the order of a solar mass but a radius comparable to the Earth radius. The forces arising from the degeneracy pressure are among the strongest in Nature, and the electron gas can keep the star stable against gravitational collapse up to the Chandrasekar limit (1.4 solar masses). The neutron stars are even more extreme, with a similar mass in a sphere of a few km

[11] By the first principle, $dE_{tot} = \delta Q + \delta L$, and here, $\delta Q = 0$.

across. In both cases, the work of Chandrasekar requires the relativistic extension of the theory.

He 3

The isotope 3 of He is produced from the decay of artificially produced Tritium; it could be a convenient fuel for fusion reactors using the reaction

$$^3He + D \rightarrow ^4 He + p,$$

where D stands for deuterium and p stands for a proton.

Here I stress that it is a Fermion and does not become superfluid like He 4. At 2.49 mK temperatures, however, the weak attractive forces produce Cooper pairs and at still lower temperatures a superfluid phase occurs.

25.6 Conjectures on Quantum Gravity

The Quantum Theory has been extended to comply with Special Relativity. The need for a relativistic formulation led Dirac to the prediction of antimatter, later verified experimentally, and Pauli to the Spin-Statistics Theorem. Dirac's theory of the electron and Quantum Field Theory give an extremely successful description of Electromagnetism, as well as Weak and Strong Interactions. Condensed Matter and Nuclear Physics are based on a firm foundation. The gravity field does not lend itself to a canonical quantization procedure like the one in use for the gauge fields. Possibly, it should not be quantized at all! At any rate, all the many attempts at a quantum theory of gravity have run into trouble so far. The existence of a quantized gravity wave, the graviton, is doubtful.[12] Roger Penrose[13] argued that the linearity of Quantum Theory is a problem, when the gravitational effects of matter are included, since General Relativity is nonlinear. The best that one can do at the present time is to write Dirac's equation and replace the derivatives by covariant derivatives, thus enabling the electron to move in a curved space–time. This approach is called semiclassical gravity theory. However, the observer introduced by Einstein to emit or receive signals at a point of space–time is inherently classical. It seems to me that the approaches that are also touted on television have a common feature: they avoid the description of an actual measurement of quantum effects, and start from the assumptions about the physics at the Planck scale, where nobody knows anything. The solar system is ridiculously small compared to the size needed for an accelerator suitable for the study of the Planck scale. As a matter of fact, there are no scientific results up to now. However, there are interesting related conjectures.

[12] "Is a Graviton Detectable?", Dyson (2013).

[13] See his book: "Fashion, Faith and Fantasy in the new Physics of the Universe", ISBN9781400880287.

25.6.1 Unruh Radiation

An interesting effect motivated by Quantum Field Theory was predicted by Stephen Fulling in 1973, by Paul Davies in 1975 and by W. G. Unruh in 1976; it is usually referred to as the Unruh effect. An accelerated observer with acceleration a in a Minkowsky vacuum should observe a thermal bath at temperature

$$T = \frac{\hbar a}{2\pi c K_B}. \tag{25.43}$$

This prediction has not been verified experimentally because in order to reach 1 ∘K, one needs an acceleration of about $10^{20} ms^{-2}$; however, the R.H.S. of the above equation is dimensionless and requires an interpretation. Moreover, putting together the zero point oscillations of the electromagnetic field (see Chap. 17), the gravitational red (or blue) shift of electromagnetic radiation and the principle of equivalence (see Chap. 8), one concludes that Eq. (25.43) is qualitatively reasonable. Above all, the result is justified by Quantum Field Theory calculations. Therefore it is taken seriously. Given the introductory level of this book, I'll present a classical argument[14] leading to Eq. 25.43. This article also contains a more advanced derivation.

Let us resume the Rindler problem (Sect. exrefsec6.2.8) of the inertial space station where a spaceship is observed to be at x at time t, while the crew proper time is τ. However we modify the flight plan. The spaceship moves along the x axis; the experiment begins when it crosses the station with a positive velocity close to c. But it has a constant negative acceleration a, which is felt like an artificial gravity by the crew. So it stops at very large x, when $\tau = 0$ (by an order of the commander, the origin of time in the ship is set from the beginning such that $\tau = 0$ at the turning point). Then the spaceship returns to the space station where it arrives with a speed approaching $-c$. Moreover, let us add a monochromatic signal emitted by the station in the x direction at frequency ω. Let us consider first what the crew sees in the first part of the trip, when the spaceship is receding from the station and start monitoring the signal. The four-vector of the radiation must be Lorentz-transformed since the observer moves at velocity given by Eq. (6.21), that is, $v(\tau) = c \tanh(\frac{a\tau}{c})$, and since $\frac{1-\tanh(x)}{\sqrt{1-\tanh(x)^2}} = e^{-x}$, using the Doppler shift formula we conclude that the crew observes a redshifted frequency that varies as $\omega e^{\frac{a\tau}{c}}$ with negative τ while receding; for the same reason it sees a frequency $\omega e^{\frac{a\tau}{c}}$ with positive τ while approaching the station. Because of these shifts the phase observed by the crew is

$$\phi(\tau) = \int_0^\tau \omega(\tau) d\tau. \tag{25.44}$$

At this point, plot twist!! Here comes the Planck formula. The spectrum is proportional to

[14] The argument (with some modification which I find to be necessary) is taken from Alsing and Milonni (2004).

$$|\int_{-\infty}^{\infty} d\tau e^{i\Omega\tau} e^{\frac{i\omega c e^{\frac{a\tau}{c}}}{a}}|^2 = \frac{2\pi c}{\Omega a} \frac{1}{e^{\frac{2\pi\Omega c}{a}} - 1}, \tag{25.45}$$

where the negative τ corresponds to the trip away from the station and the positive τ to the return.

The simple argument is not entirely convincing but is certainly suggestive of a relation between the Doppler effect, acceleration and the Planck formula. On the other hand, we know from Sect. 17.6 that the vacuum contains the zero point fluctuations of the field and it is reasonable that the vacuum calculated with Rindler coordinates is different, so it has different modes.

25.6.2 Black Hole Evaporation?

There is hardly anything more unbalanced and out of equilibrium than a black hole, which is a catastrophic, ongoing gravitational collapse that we see strongly redshifted by gravity. After an *infinite time*, however, nothing changes any more. Stephen Hawking and many authors refer to black holes but not the real swirling and catastrophic ones: actually, they consider the peaceful so-called eternal black hole, identified with the Schwarzschild vacuum solution with point mass (the so-called singularity) and horizon.

Since nothing comes out of a black hole, the total area of their event horizons should never shrink. For instance, if two black holes collide and merge, the area of the resultant black hole must exceed the sum of the areas of the two colliding holes. This conclusion, which is not hard to believe, is called Hawking's area theorem (1971). The study of gravitational waves emitted in the collision of black holes confirmed this result.[15] In 1972, Jacob Bekenstein noted that since the black holes can only get bigger and the entropy of an isolated system can only increase there is an analogy.

Hawking took the analogy seriously. If the black hole has an entropy, assuming equilibrium it must also possess a temperature. A temperature implies the exchange of heat with a thermal bath, and the emission of thermal radiation contradicts the very definition of a (eternal) black hole. The original development requires quantum field theory in curved space time, is involved and requires approximations and assumptions. Some alternative derivations use the analogy with the Unruh effect, trying to show that a close observer, by the equivalence principle, should see a strong Unruh radiation and we should see some of it from the Earth. Unfortunately, if we are far from the hole, the acceleration here is zero, the Unruh effect is nothing, and attempts to justify the transfer of the observed radiation from the observer near the horizon to the ground are not clear. Other derivations (also by Hawking) pretend to show that near the horizon there is a pair production with positive energy particles (somehow) coming out and negative energy ones preferentially going down in order

[15] Isi, et al. (2021).

to decrease (somehow) the black hole mass. None of the derivations that have been given for the prediction of evaporation is fully convincing, the various derivations are not mutually consistent, and quantum gravity corrections could be important.[16]

This is an introductory book and I come to the original conclusions reported in the literature. A Schwarzschild black hole with mass M emits a black body radiation at temperature

$$T = \frac{\hbar c^3}{8\pi G M K_B}. \tag{25.46}$$

Equation (25.46) is striking and celebrated, because it is the first to contain both G and $\hbar$, and so it smells like a start of the unified theory which is greatly desired, particularly by the believers in quantum gravity. In this way a perfectly isolated black hole should lose mass by radiating a power

$$P = 3.5\frac{10^{32}}{M^2}\,\mathrm{W}, \tag{25.47}$$

where M is in Kg, and it should evaporate in a time

$$\sim 8.410^{-17} * M^3\mathrm{s}. \tag{25.48}$$

The process should start very slowly, but speed up and produce a final bang of X-Rays. So the black hole would not be the final fate of the matter falling in, which should be simply waiting for a resurrection in the form of thermal photons, but after a very, very long time. For a solar mass ($2\ 10^{30}$ Kg) the evaporation time has been evaluated as 10^{67} years, but its present temperature should actually be much lower than the background temperature of the Universe, and so it should continue growing for an absurdly long time.[17]

If the Big Bang at the origin of the Universe produced black holes of the right mass, similar to a mountain, it would be so kind of them to produce their own small bang right now, thus giving an experimental support to this theory. No observational confirmation of evaporating black holes has been obtained so far, and by the way, we also know that the current theory cannot be fully correct.

If one believes that the conclusion (25.46) is correct, it says that most probably there are no observable consequences of the theory. However, there could be a Hawking radiation if there are quantum corrections ignored by the theory. Also, people working in quantum gravity discuss the possibility that black hole evaporation produces a loss of information (information paradox, which, however, is questionable and has no observable consequences).

---

[16] See the paper by Helfer (2003).

[17] In mathematics, 1 is an integer and 10^{10000} is just another integer. Physics is not mathematics. The proton is known to be stable for more than 10^{34} years, but theories covering immensely longer times have no meaning and are of no interest for Science, since the scientific method requires experiments.

25.7 Hanbury Brown and Twiss Effect: Bosons and Fermions

One speaks about Hanbury Brown and Twiss effect when the correlation in the simultaneous detection of pairs of identical particles shows that the particles follow a non-classical statistics. The original discovery[18] was done with photons. The experiment was aimed at the measurement of the angular radius of Sirius, which turned out to be 3 billionths of a radiant. In Sect. 4.5.1, a classical account of the effect is briefly presented, and it is shown that the effect is borne out by the Maxwell equations. It is also interesting to discuss the effect from the quantum viewpoint, in terms of the photons that obey Bose statistics. While the simultaneous arrival of pairs of photons, one in each detector, was not measured, the correlation function (4.19) represents an average thereof, and would have been constant if the photons arrived at random. Since different photons are emitted from different atoms, say, 1,000,000 km apart, somebody could expect that the photons forget about the quantum statistics and the permutation symmetry. However, this is experimentally not the case! As pointed out by Fano, the correct reasoning is in terms of pairs of photons. If two atoms a and b emit two photons that are collected by detectors 1 and 2, there are two amplitudes: $A(a \to 1, b \to 2)$ and $A(a \to 2, b \to 1)$; the probability is $A(a \to 1, b \to 2) \pm A(a \to 2, b \to 1)$, with the upper sign holding for Bosons and the lower sign for Fermions.

A direct experimental verification of the above ideas in the laboratory was made possible[19] by the use of ultra-cold He* atoms at 0.5 μ°K. Here, He* is a symbol for excited atoms in a 1s2s (singlet or triplet) configurations; such configurations are metastable, because the 2 s→1 s decay is strictly forbidden with one photon.[20] The atoms were prepared and dropped at a macroscopic height L above a detector, where they fell by gravity. The detector allowed for the construction of the normalized correlation function $g^{(2)}(\Delta r)$, i.e., the probability of joint detection at two points separated by Δr, divided by the product of the single detection probabilities at each point. Statistically independent detection events would give $g^{(2)}(\Delta r) = 1$. A value larger than 1 indicates bunching, that is, the Bose-like tendency to have a larger-than-classical population of the phase-space cell; a value less than 1 is evidence of Fermion-like anti-bunching. The experiment was done with ^{4}He* (bosons) and ^{3}He* (fermions), showing that $g^{(2)}(\Delta r) = 1$ beyond a correlation length on the order of 1mm, with a clear increase at short distances for Bosons and a decrease for Fermions. The correlation length was $l = \frac{\hbar t}{ms}$, where t is the falling time, m the mass and s the vertical size of the source.

The anti-bunching correlation has been experimentally observed by Kiesel et al. in 2002 and Kimble et al. for Na atoms in 1977. A Fermionic Hanbury Brown and

[18] Brown and Twiss (1956).

[19] "Comparison of the Hanbury Brown and Twiss effect for boson and Fermions", by T. Jeltes et al., http://archiv.org/abs/cond-mat/0612278.

[20] The photon is odd, so the transition violates Parity. It takes place with two photons, but auto-ionization (similar to the Auger effect) is more probable.

Twiss experiment[21] has been performed at 2.5 °K by partitioning an electron beam by means of a metallic gate in a two-dimensional electron gas in the quantum Hall effect regime. The fluctuations in the currents in the two partial beams were found to be anti-correlated, as expected.

References

Alsing, P.M., Milonni, P.: arXiv:quant-ph/0401170v2 (2004)
Brown, R.H., Twiss, R.Q.: A test of a new type of stellar interferometer on Sirius. Nature **178**, 1046–1048 (1956)
Cini, M.: Solid State Commun. **20**, 695 (1976)
Cini, M.: Solid State Commun. **24**, 681 (1977)
Dyson, F.: Int. J. Mod. Phys. A **28**, 1330041 (2013)
Helfer, A.D.: arXiv:gr-qc/0304.0042v1 (9 Apr 2003)
Henny, M., et al.: Science **284**, 296 (1999)
Isi, M., et al.: Phys. Rev. Lett. **127**, 011103 (2021)
Kanamori, J.: Prog. Theor. Phys. **30**, 275 (1963)
Sawatzky, G.A.: Phys. Rev. Lett. **39**, 504 (1977)

[21] Henny, et al. (1999).

Chapter 26
Quantum Transport and Quantum Pumping

Quantum transport deals with many-body and off-equilibrium problems and is a very active topic of research.

26.1 Fundamental Quanta of Resistance

In 1826, Georg Simon Ohm published the basic law $V = RI$ of linear circuits, where V is the potential difference between the ends of a metallic wire, I the current and R the resistance. For a cylindrical wire, $R = \rho\frac{L}{A}$, where L is the length of the wire, A the area of its cross section and ρ is a material property called resistivity, while $\sigma = \frac{1}{\rho}$ is the conductivity. The old, classical Drude theory which we have already talked about in Chap. 4 treats the conduction electrons like point charges which are accelerated by the field but scatter against the ion cores of the metal, producing a sort of friction. This theory is qualitatively correct for Alkali metals, but predicts that for small L the resistance should vanish and current should diverge. Actually, the high school laws of the circuits are only valid in the macroscopic world. However, when the linear dimensions are less than an electron mean free path λ, which may be on the order of 10 nm (22 in Al and 55 in Cu), the motion of the electrons is ballistic, that is, the circuit obeys quantum laws. There is a lot of interest in nano-circuits. The atoms form a lattice, and the electrons can diffract unhindered through a perfect periodic potential, but scatter against phonons, impurities, lattice defects and walls. Scattering events dissipate energy and produce Joule heat.

It must be stressed that in the quantum world, the dimensionality has a much stronger meaning than in classical Physics. In thin films that are extended in the $x - y$ plane, there are excitations along z, but if the thickness is L_z, the creation of such excitations takes energies on the order of $\frac{\hbar}{2mL_z^2}$, where m is the electron mass; therefore,

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_26

the low-energy and low-temperature Physics is almost completely two-dimensional. In the same way, nano-wires and Carbon Nano-tubes can be really well described by one-dimensional models. The theoretical analysis of such systems is made more difficult by an enhanced importance of the interactions. In low-capacitance tunnel junctions under small bias, one observes the *Coulomb blockade*: the presence of an electron in a Quantum Dot prevents other electrons from jumping, and one observes an increased resistance. Needless to say, nano-devices are extremely fashionable today in science and technology.

26.1.1 Quantum Transport in Nanoscopic Devices

For nanoscopic objects (smaller than the electron mean free path λ), one can ignore energy dissipation, but the transport is *ballistic*, i.e., well described by the electron wave functions and Green's function techniques (see next section). However, one can easily understand that the resistance must be quantized.

Consider a junction such that the Fermi level makes an abrupt jump eV; the current is $I = e\nu$, where ν is the hopping frequency. The energy jump corresponds to the frequency $\frac{eV}{h}$; including the spin degeneracy we find $\nu = 2\frac{eV}{h}$, so $I = 2\frac{e^2V}{h}$. Hence, the quantum of conductance is $G = \frac{2e^2}{h}$ and the quantum of resistance is

$$R = \frac{h}{2e^2}. \tag{26.1}$$

This is $\sim$ 12900 Ohm. In macroscopic conductors, there are normally many channels and the resistance can be much lower. In reality, any nanoscopic conductor behaves like a bunch of M conducting channels; for a two-dimensional device, one has $M \sim WK_F$ conducting channels, where W is the width and K_F the wave vector of an electron at the Fermi level. Typically, M is of the order of 10 in nano-wires. Experimentally[1] by applying a negative gate potential across the sample, one can switch on and off the conducting channels. In this way, the quantization is evident.

The resistance quantum has a special meaning for nano-circuits where quantum effects dominate. It shows up in the Quantum Hall Effect, discovered by Klaus von Klitzing which will be discussed below, and allows to measure R very accurately. This is important for fundamental Physics because it is related to the fine structure constant

$$\alpha = \frac{2\pi e^2}{hc} \sim \frac{1}{137.035999084\cdots}.$$

[1] van Wees et al. (1988).

26.1.2 *Magnetoresistance, Giant Magnetoresistance and Colossal Magnetoresistance*

Some materials have a resistivity that depends on an applied magnetic field. This effect is called Magnetoresistance and was discovered by Lord Kelvin in 1856. Since the resistance is due to the scattering of conduction electrons close to the Fermi level and the same electrons also conduct electricity, the very fact that the magnetic properties affect the resistivity is rather obvious. Most materials show only a change by a few percent with the feasible magnetic fields. However, the effect is large in some materials. The so-called Colossal Magnetorsistance (CMR) was discovered by Jonker and Van Santen in the 1950s in Manganese-based perovskite oxides and implies resistance changes by orders of magnitude; the alignment of the Mn spins by the field produces enhanced conductivity. This finds application in hard disks.

The 2007 Nobel prize was awarded to Albert Fert and Peter Grünberg for the discovery of the Giant Magnetoresistance. They discovered important changes (up to 50%) in the resistance of materials in which alternating layers of different transition metals like Fe and Cr. If the spin orientation of adjacent layers is parallel the resistance is low, while for anti-parallel orientation it becomes higher. This has a simple qualitative explanation since conduction electrons with spin parallel to the magnetization of the layers have low resistance, while electrons with antiparallel spin have high resistance and the mixing of the two populations is not important.

This effect produces a change in resistivity by some percent, which is important and is exploited in magnetic field sensors.

26.2 Characteristics and Transients

In the original treatment,[2] I used the Keldysh formalism, but here I propose a much simpler presentation for an independent particle model. Let $c_m^\dagger$ denote the electron creation operator at site m in an infinite discrete model including the electrodes; spin is understood. The operator for the current is (Eq. (22.21)

$$J_m = \frac{e\tau_{mn}}{i\hbar}(c_m^\dagger c_{m-1} - c_{m-1}^\dagger c_m). \tag{26.2}$$

In a preliminary calculation, suppose that the system is in an eigenstate of the initial Hamiltonian H_i which must describe a device connected to both left and right electrodes, with zero potential difference. Suppose that initially, at time $t = 0$, the system is not biased and the many-electron ground state wave function is a determinant of orthonormal spin-orbitals $|q\rangle$, each labeled by some index q which might be an asymptotic wave vector, say, in the right electrode (Figs. 26.1 and 26.2).

[2] Cini (1980), (2014).

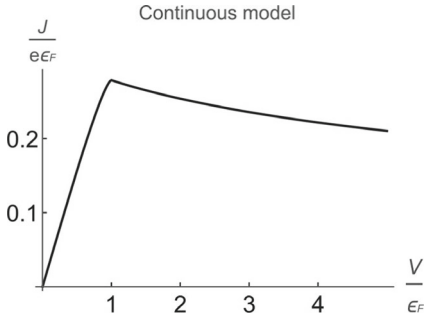


Fig. 26.1 Current–Voltage characteristic of the one-dimensional continuous model

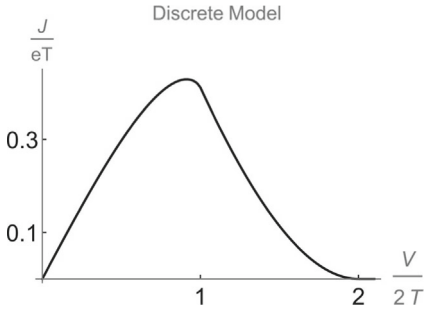


Fig. 26.2 Current–Voltage characteristic of the one-dimensional continuous model

How does the determinant evolve when the bias is on? At time t, the one-electron state $|q\rangle$ becomes $|q(t)\rangle$ and the initial determinant becomes the determinant of evolved one-particle states. Are the one-particle states still orthonormal in the evolved system? The answer is yes. The evolution operator $U(t)$ is unitary, so it preserves normalization and orthogonality.

Now we know how to evolve a large many-electron determinant, by evolving individual spin-orbitals, but we must cope with thermal equilibrium. In equilibrium at some temperature T, the one-body states $|q\rangle$ are occupied with probability given by the Fermi distribution $f_q(T)$ and the average current vanishes. Let us switch the bias on at time $t = 0$; the system starts evolving with a final-state Hamiltonian $H_f(t)$, and the current at site m at time t is given by

$$\langle J_m(t)\rangle = \sum_q f_q \langle q|U^\dagger(t) J_{mn} U(t)|q\rangle. \tag{26.3}$$

We can also put the Fermi distribution inside the matrix element, as an operator:

$$\langle J_m(t)\rangle = \sum_q \langle q|U^\dagger(t) J_m U(t)\hat{f}|q\rangle, \tag{26.4}$$

with $\hat{f} = (\exp(\beta(H_0 - \mu) + 1)$ and with H_0 the unbiased Hamiltonian. The states that were occupied before the switching of the bias are those that carry the current

at all times. Thus, the exact current through site m in a discrete model with hopping parameter t_h is given by

$$\langle J_m \rangle = \frac{2et_h}{\hbar} \mathrm{Im} \sum_q f_q g^r_{m,q}(t) g^r_{m-1,q}(t)^*, \tag{26.5}$$

where the retarded Green's function is given by

$$i g^r_{m,j}(t) = \theta(t) \langle j | U(t) | i \rangle, \tag{26.6}$$

and $|i\rangle$ is the one-electron wave function localized at site i. In this way, the problem reduces to a one-body time-dependent calculation.

It is instructive to consider the case when the initial Hamiltonian $H = H_i$ represents an infinite wire (in a discrete or continuous model), while at time $t = 0$ the Hamiltonian jumps to $H_f = H_i + \hat{V}$ where $\hat{V}$ is a potential that raises the left-hand half of the wire by V. Then, for $t > 0$, $U(t) = \exp(-iH_f t)$.

In this way, one can solve, for instance, for the transient current following the onset of the bias; also, letting $t \to \infty$, one can use the asymptotic expansion techniques to derive the dependence of J on V in the steady state; $J(V)$ is called the current–voltage characteristics of the model. A similar treatment is also available for continuous models. In Fig. 26.1, we see that the current is about linear for small V but it changes slope suddenly and then it starts decreasing. This is evidently an effect of dimensionality.

Besides, this formalism allows us to follow the transient effects after the bias is switched off; in the above examples the current starts from zero, then overshoots the asymptotic $t \to \infty$, oscillates around it a few times and converges to the final value with a characteristic time given by the inverse bandwidth.

Such results were the first time-dependent exact treatment of quantum transport in 1980, but now by the use of the Keldysh formalism and numerical techniques the interaction effects are also included. Currently many groups are engaged in developing codes to deal with electron–electron interaction effects in realistic situations. The development of magnetic and relativistic effects has not received much attention to date, but will be a challenge in the near future.

26.2.1 Landauer Formula

In 1957 Landauer studied the current J through a quantum system, like a thin film, connected to leads as a function of the potential difference V, namely the current–voltage characteristic $J(V)$. He proposed the following, phenomenological, smart and almost self-explanatory formula for the current–voltage characteristic:

$$J(V) = \frac{e}{\hbar} \int d\epsilon [f_L(\epsilon) - f_R(\epsilon)] \rho^{(L)}(\epsilon) \rho^{(R)}(\epsilon) \sum_{m,n} |2\pi V_m^R V_n^{L*} g^{(r)}_{m,n}(\epsilon)|^2. \quad (26.7)$$

Here, f_L and f_R are the Fermi functions of the left and right electrode, $\rho^{(L)}$ and $\rho^{(R)}$ are the densities of states, $g^{(r)}$ is the retarded Green's function (see the previous section) giving the amplitude of incoming in eigenstate m of left electrode and outgoing in state n or right electrode. A multi-terminal generalization was provided by Büttiker.

26.3 Quantum Pumping, Berry Phase and Chern Numbers

Quantum Pumping is the generation of current across a circuit by acting locally on some parts of it, without applying a ddp between the ends; several mechanisms have been discovered so far by theorists, starting with a seminal paper[3] by the Nobel prize Thouless.[4] He considered an elementary one-dimensional insulator model with Hamiltonian $H_0 = \frac{p^2}{2m} + V(x)$, with $V(x + a) = V(x)$; the Bloch functions of band n satisfy Bloch's theorem $\psi_{nq}(x) = e^{iqx} u_{nq}(x)$. The periodic function $u_{nq}(x)$ satisfies $H_0](q)u_{nq}(x) = \epsilon_{nq} u_{nq}(x)$ with $H(q) = e^{-iqx} H e^{iqx}$. The Fermi level is assumed to lay in the gap between bands n and $n + 1$.

Suppose we add a weak time-dependent perturbation with the same spatial periodicity a added to H_0. The perturbation is so mild that the instantaneous Hamiltonian keeps the Fermi level within the gap. At first sight, no time dependence can break the left–right symmetry and no current can be excited; remarkably, the first sight is misleading. Now the total Hamiltonian is also periodic in time with period T: $H(q, t) = H(q, t + T)$ holds, and $H(q + \frac{2\pi}{a}, t) = H(q, t)$. The parameter space is a torus. The Bloch states are of the form $e^{iqx} u_n(q, t)$, and since there are two parameters there are a Berry connection

$$A_n = -i(\langle u_n | \frac{\partial u_n}{\partial q} \rangle, \langle u_n | \frac{\partial u_n}{\partial t} \rangle) \quad (26.8)$$

and a curvature

$$\Omega^n_{qt} = -i[\langle \frac{\partial u_n}{\partial q} | \frac{\partial u_n}{\partial t} \rangle - \langle \frac{\partial u_n}{\partial t} | \frac{\partial u_n}{\partial q} \rangle] \quad (26.9)$$

which represents a drift velocity of the electrons. Thouless by adiabatic perturbation theory shows that the velocity of band electrons is $v_n(q) = \frac{1}{\hbar} \frac{\partial \epsilon_n}{\partial q} + \Omega^n_{qt}$, and while the average of the first term over a filled band vanishes (the electrons go to left and to right) the drift velocity produces a pumped current

[3] Thouless (1983).

[4] David Thouless, British, 1934–2019, Nobel prize in 2016 with F. Duncan and J. M. Kosterlitz.

$$j = -e \sum_n \int_{BZ} \frac{adq}{2\pi} \Omega_{qt}^n, \tag{26.10}$$

which implies that the number of pumped electrons per cycle is

$$c_n = -\int_0^T dt \frac{j}{e} = -\int_0^T dt \int_{BZ} \frac{dq}{2\pi} \Omega_{qt}^n. \tag{26.11}$$

Note that $2\pi c_n$ is the flux of the Berry *magnetic field*, that is, a Berry phase. Hence,

$$-2\pi c_n = \int_A^B dx A_x(x, 0) + \int_B^C dy A_y(1, y) + \int_C^D dx A_x(x, 1) + \int_D^A dx A_y(0, x), \tag{26.12}$$

where A = (0,0), B = (1,0), C = (1,1) and D = (0,1). There are matching relations at opposite sides of the square, $u(x, 0) = e^{-i\theta_x(x)}$, $u(x, 1)$, $u(0, y) = e^{-i\theta_y(y)}$ and so on.

From this calculation, it turns out that $c_n = \theta_x(1) - \theta_x(0) - \theta_y(1) + \theta_y(0)$ must be an integer (as it is expected, since it is the number of electrons). Such integers which result from Berry phase calculations are called Chern numbers[5]; the above calculation is an example of the way integers arise from Berry phases.

Several different mechanisms were invented later for sending a charge current or even a spin current by pumping. Brower[6] has shown that there is a connection between a class of pumping mechanisms, based on varying two parameters in the Hamiltonian, and the Berry phase. However there are also non-adiabatic mechanisms of pumping. An example[7] is shown in Fig. 26.3. The model is a simple discrete ring laterally connected to a wire. As shown in the figure, a laterally connected ring carrying a counterclockwise current will pump charge to the right. For each flux quantum swallowed by the ring, there is an amount of charge shifted from left to right. However this amount decreases if the flux is inserted slowly, and this shows the non-adiabatic character of the process. Many groups are engaged in the research of spin-polarized quantum pumping that one day could replace electronics with spintronics.

26.4 Hint to Spontaneous Symmetry Breaking and Goldstone Theorem

Imagine a system that has a certain symmetry, meaning the laws of physics are the same when you perform a transformation on it (like a rotation). In some cases, the ground state (or vacuum) of the system doesn't respect this symmetry, even though

[5] After the great Chinese mathematician Shing-Shen Chern (1911–2004).

[6] Brower (1998).

[7] Cini and Perfetto (2011).

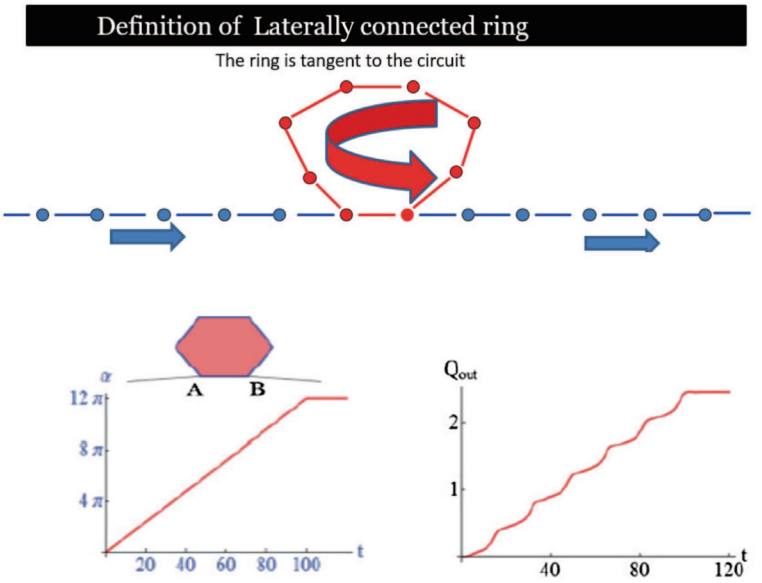


Fig. 26.3 Top: in a laterally connected ring, the choice of a chirality also defines a direction along the wire. Bottom left: a hexagonal ring and the flux piercing it as a function of time. The time is measured in units of $\frac{\hbar}{\tau}$, where τ is the matrix element between neighboring sites. Bottom right: charge pumped from left to right in the wire in the process; each fluxon through the ring produces a step

the underlying laws do. A classical example is provided by a pencil balanced perfectly upright on its tip; it can fall in any direction, breaking the rotational symmetry.

Spontaneous symmetry breaking occurs when the ground state of the system chooses a specific configuration that breaks the original symmetry, even though the equations describing the system are symmetric. The Goldstone theorem is a fundamental concept in quantum field theory that is related to spontaneous symmetry breaking. It was formulated by Jeffrey Goldstone in 1961 and plays a crucial role in understanding certain aspects of particle physics, particularly in the context of the Higgs mechanism and the emergence of mass in gauge theories. In the next section about Superconductivity, we shall learn how photons can acquire mass in this way. The Goldstone theorem states that if a continuous symmetry is spontaneously broken in a quantum field theory, then there must exist massless scalar particles called Goldstone bosons. These bosons are associated with the broken symmetries and are responsible for the long-range forces associated with the broken symmetry. Any solid breaks the perfect translational symmetry of space by choosing lattice sites for the atoms; the translational symmetry is reduced to the lattice translations. But then the lattice vibrations arise, and these are the Goldstone bosons. The Higgs mechanism

in the Standard Model of particle physics is an example where the Goldstone theorem plays a role. The Higgs boson itself arises from the spontaneous breaking of electroweak symmetry and interacts with other particles to give them mass.

26.5 Superconductivity

In 1911, Kamerlingh Onnes discovered superconductivity, and if I say that the discovery was unexpected, this is an understatement. It was a mystery, indeed something unbelievable, and seemed contrary to the common sense and the idea that in the real world nothing is perfect. The discovery stunned everyone; it could just be reproduced at will by measuring the resistance of a piece of Hg versus temperature. Below a critical temperature T_C, which in the older *conventional* superconductors is no more than a few degrees Kelvin, the electrical resistance is nothing. Please believe me, I do not mean that it is small; it is just 0. Moreover, the superconductor is a perfect diamagnet, expelling the entire magnetic field from the bulk. If a piece of superconductor is put in a field, this is allowed within a thin penetration depth, and repelled. The repulsion levitates objects. Magnetically levitated trains can travel at 500 Km per hour. This phenomenon is one of the first discovered examples of the *magic of Quantum Mechanics,* but several decades passed before it was understood. It is very important in Condensed Matter Physics where superconductivity was discovered, but it is quite relevant in Nuclear Physics, stellar Astrophysics and Theoretical Physics in general. For these reasons it deserves to be highlighted here.[8]

In superconductivity, most of the electrons behave as they do in ordinary paramagnetic metals, but a strict minority of them forms bound pairs, thus producing a drastic change of the macroscopic properties; the most important changes are zero resistance and perfect diamagnetism. In the superconducting phase a gap $E_{gap} = 2\Delta$ is observed in the single particle spectrum (e.g., in Photoemission experiments); the gap implies that in order to excite an electron, one needs an energy exceeding E_{gap}. In a normal metal, the electrons fill the states below the Fermi level, while in superconductors there is a forbidden gap, typically $E_{gap} \sim 3.5 K_B T_C$ (actually $E_{gap} \sim 4.1 K_B T_C$ for Pb and $E_{gap} \sim 3.3 K_B T_C$ for Cd). So,T_C depends on the gap. For those metals that become superconductors, the critical temperature is a fraction of Kelvin (0.39 for Ti) to several degrees (6.55 for La,9.26 for Nb).

For normal metals, the current depends linearly on the electric field (Ohm's law reads $\vec{J} = \sigma \vec{E}$); in a superconductor the electric field accelerates the electrons,

[8] In 1935, J. N. Rjabinin and Lev Shubnikov made another unexpected experimental discovery, namely type II superconductors, like Niobium and Niobium–Titanium alloys. These have a superconducting phase, but above a critical combination of temperature and magnetic field, they develop a mixed phase with some penetration of the magnetic field in the form of vortices. The theory of type II superconductors was developed by Abrikosov. However in this book we consider only type I superconductors.

there is no resistance, and it holds that $\frac{\partial(mv)}{\partial t} = -e\overrightarrow{E} \equiv \frac{\partial \overrightarrow{J}}{\partial t} = \frac{ne^2}{m}\overrightarrow{E}$; the rotor gives $rot\frac{\partial \overrightarrow{J}}{\partial t} = \frac{ne^2}{m} rot\overrightarrow{E}$ which, using the Maxwell equations, gives

$$-\frac{ne^2}{mc}\frac{\partial \overrightarrow{B}}{\partial t} \equiv \frac{\partial}{\partial t}(rot\overrightarrow{J} + \frac{ne^2}{mc}\overrightarrow{B}) = 0.$$

But in a superconductor a stronger condition holds. It is the London equation

$$rot\overrightarrow{J} + \frac{ne^2}{mc}\overrightarrow{B} = 0. \tag{26.13}$$

In the gauge $div\overrightarrow{A} = 0$, one can write

$$\overrightarrow{J} = -\frac{ne^2}{mc}\overrightarrow{A}(\overrightarrow{x}). \tag{26.14}$$

For a stationary field $rot\overrightarrow{B} = \frac{4\pi}{c}\overrightarrow{J} \rightarrow rot(rot\overrightarrow{B}) = \frac{4\pi}{c} rot\overrightarrow{J}$, but since $rotrot = graddiv - \nabla^2$,

$$\nabla^2\overrightarrow{B} = \frac{4\pi ne^2}{mc^2}\overrightarrow{B} \tag{26.15}$$

$$\nabla^2\overrightarrow{J} = \frac{4\pi ne^2}{mc^2}\overrightarrow{J}. \tag{26.16}$$

Now one can solve the boundary value problem of a surface exposed to an external magnetic field, and find that the field is exponentially damped inside, with the London penetration depth

$$\Lambda = \sqrt{\frac{mc^2}{4\pi ne^2}} = \frac{c}{\omega_p}. \tag{26.17}$$

Remarkably, from the above equations it turns out that in the preferred gauge which is most natural to describe superconductivity (but corresponds to the loss of a degree of freedom), the vector potential also becomes short-ranged, as if the photons had gained a mass. This is considered to be a simple forerunner of the Anderson–Higgs mechanism whereby the Higgs boson gains mass by a spontaneously broken symmetry.

26.5.1 Flux Quantization

Consider a ring and an electron therein with a wave function $\psi(\overrightarrow{x})$. We can introduce a vector potential $\overrightarrow{A} = \nabla\chi(\overrightarrow{x})$ with a gauge transformation $\psi(\overrightarrow{x}) \rightarrow$

$\psi(\overrightarrow{x})\exp(\frac{e\chi(\overrightarrow{x})}{c\hbar})$. We can pick $\chi(\overrightarrow{x}) = \int_{\overrightarrow{x}_0}^{\overrightarrow{x}} \overrightarrow{A}\cdot d\overrightarrow{r}$, where $\overrightarrow{x}_0$ is an arbitrary point of the ring. Now suppose that there is no magnetic field in the metal ring, but there is a flux tube in the hole. Then $\oint \overrightarrow{A}\cdot d\overrightarrow{r} = \Phi$ is the magnetic flux. But the wave function must have a single value. This requires that in a tour of the circumference $\frac{ie}{c\hbar}\Delta\chi = \frac{ie}{c\hbar}\Phi = 2\pi mi$ with integer m. This requires

$$\Phi = m\phi_0 \tag{26.18}$$

where

$$\phi_0 = \frac{hc}{e} \tag{26.19}$$

is the flux quantum. The ground state wave function and energy of the ring are periodic functions of the flux, and therefore the flux is quantized. In the case of superconductors, the experiment shows that the above is true but is not the whole truth, because the flux is quantized in units of $\frac{\phi_0}{2}$.

26.5.2 Clues for Pairing

Hence, Onsager suggested in 1953 that in superconductors the charge carriers are not electrons but pairs of electrons. These must have appeared lunatic, since the electrons repel each other. The screened Coulomb interaction in metals $V_s(q) = \frac{4\pi e^2}{\epsilon(q,\omega)q^2}$ is still repulsive and does not explain any pairing, but another crucial hint (after the one about flux) came from experiments: there is an isotope effect. The transition temperature is inversely proportional to the square root of the mass of the isotope. Then, Fröhlich suggested that the electron–phonon interaction is involved. Indeed, each electron interacts with the phonons and tends to surround itself with positive charges, which can be attractive for a second electron having the opposite wave vector (and hence, the same periodicity). Thus, to second-order in the electron–phonon interaction, the electrons experience an attractive, pairing interaction.

26.5.3 Superconductivity Theory: the Line of Reasoning

Superconductivity is a general phenomenon that manifests itself under different conditions with disparate mechanisms. It is believed that even in neutron stars, there is superconductivity. The explanation of superconductivity has been under discussion for many years as far as cuprates are concerned. Below I give a simplified account of the most simple form of the theory, which is considered valid at least for the conventional low-temperature superconductors like Pb. My experience with the students is that many fail to grasp the physics even after working out all the algebra.

Therefore, before developing mathematics, it is useful to clarify the logical line we intend to follow and it is quite intricate. First of all, following Cooper, we have to show that assuming that there is attraction it makes sense to talk about bound pairs of electrons. To this end, we assume an electron gas in which all the electrons are in the ground state of the non-interacting Fermi distribution (Fermi sphere); then we consider adding a pair of extra opposite spin electrons with an attractive pairing interaction and find that the Pauli principle allows for the formation of a bound pair.

Good, in this way we have shown that our starting assumption is wrong, since if there is pairing then all the other electrons should be paired and should not form a Fermi sphere. Are we lost?

No. The reason is that in real superconductors, the kinetic energy dominates, and almost all the electrons do indeed form the Fermi sphere. So, only relatively very few close to the Fermi energy do pair. In this way, the Cooper pair is a reasonable hint to what happens. This is developed in the next paragraph. In the following paragraph, nevertheless, we must find how these coupled pairs are organized; to this end, we develop the so-called BCS theory. It turns out that the pairs form a special many-body state which has strong analogy with a coherent many-boson state.

The final paragraphs of this section are devoted to the physical consequences of all that, and in particular to the explanation of the fact that having a coherent state of pairs (see Chap. 16) leads to properties like zero resistance and perfect diamagnetism. Now we are ready to begin.

26.5.4 Cooper Pairs

Cooper in 1956[9] considered the formation of a bound pair of electrons in the presence of a Fermi sphere of occupied states with wave vectors below the Fermi level (i.e., $k \leq k_F$, where k_F is the Fermi wave vector). So, the electron band is represented by the unperturbed Hamiltonian

$$H_0 = \sum_{\overrightarrow{k},\sigma} \epsilon_k n_{k,\sigma}. \tag{26.20}$$

We include the phonons by the simple Debye model, in which the phonon frequencies go from zero to a maximum given by the Debye frequency

$$\omega_D = q_D v_s, \tag{26.21}$$

where q_D is the maximum phonon wave vector, of the order of the inverse lattice constant, and v_S stands for the velocity of sound in the crystal. The presence of the phonons implies that there are electrons with energies between the Fermi energy and

[9] Physical Review. 104 (4): 1189–1190.

the Fermi energy plus $\hbar\omega_D$. The added Cooper pair is assumed to be in this energy range. So, we add a pair having opposite wave vectors and spins, in a state

$$|d[k]\rangle = |k_\uparrow, -k_\downarrow\rangle. \tag{26.22}$$

The interaction is also provided by the phonons.

The effect of the interaction is introduced through the perturbation

$$H' = -V \sum_{\vec{k}\,,\,\vec{k'}\in g} |d[k']\rangle\langle d[k]|, \tag{26.23}$$

where $V > 0$ and g is defined by $k \in g \equiv E_F \leq \epsilon_k \leq E_F + \hbar\omega_D$.

Having set the scene this way, our task is to determine what happens with this pair. We look for solutions of $H|\Phi\rangle = E|\Phi\rangle$, of the form

$$|\Phi\rangle = \sum_k \phi_k |d[k]\rangle \tag{26.24}$$

knowing that $H_0|\Phi\rangle = \sum_k 2\epsilon_k\phi_k|d[k]\rangle$, while

$$H'|\Phi\rangle = -V \sum_{k,k'} |d[k']\rangle\langle d[k]| \sum_{k''} \phi_{k''}|d[k'']\rangle. \tag{26.25}$$

An exchange of dummy indices yields

$$H'|\Phi\rangle = -V \sum_{k,k'} \phi_{k'}|d[k]\rangle.$$

Hence, we obtain

$$(2\epsilon_k - E)\phi_k = V \sum_{k'} \phi_{k'}.$$

The R.H.S. is a constant N, and we may write

$$\phi_k = \frac{N}{2\epsilon_k - E}. \tag{26.26}$$

The eigenvalue equation is

$$1 = V \sum_k \frac{1}{2\epsilon_k - E}. \tag{26.27}$$

Introducing a density of states per spin ρ, we may write

$$\frac{1}{V} = \int_{E_F}^{E_F+\omega_D} \frac{d\epsilon}{2\epsilon - E}. \tag{26.28}$$

Since ρ can be considered as a constant over the narrow interval of integration, one readily arrives at a shift of E from the unperturbed value $2\epsilon_k$ to $2\epsilon_k - \Delta$; the binding energy is

$$\Delta = \frac{2\omega_D}{e^{\frac{2}{V\rho(E_F)}} - 1}. \tag{26.29}$$

Therefore,[10] even for small V, the pair gets bound ($\Delta > 0$).

An added pair gets bound. We conclude that the Fermi sphere is unstable in the presence of a pairing interaction; for the same reason, our starting assumption was wrong. But now we have the hints for a satisfactory theory.

26.5.5 The Bardeen–Cooper–Schrieffer (BCS) Hamiltonian

The above results by Cooper imply that, in the presence of an attractive electron–electron interaction, the Fermi sphere becomes unstable. To explain superconductivity, we need to find the ground state formed by the pairs and also the corresponding low-energy excitations that determine the amazing physical properties of superconductors. The accepted theory is the one due to John Bardeen, Leon N. Cooper and John Robert Schrieffer[11] (BCS) (Nobel prize in 1972).[12] The ingredients of the BCS theory are bare electrons, the Coulomb repulsion, the phonons and the electron–phonon interaction. BCS show that the electrons close to the Fermi level with opposite spin and momentum form bound pairs and produce the observed physical effects.

Here we wish to consider how the principle works, and, following BCS, we deal with a simplified model, or reduced Hamiltonian, which starts from the Fermi sphere plus an attractive interaction among electrons having opposite wave vectors. Here is the model, in obvious notation:

$$H_{BCS} = \sum_{\mathbf{k}} \epsilon_{\mathbf{k}} (n_{\mathbf{k}\uparrow} + n_{-\mathbf{k}\downarrow}) + \sum_{\mathbf{k},\mathbf{k}'} V_{\mathbf{k},\mathbf{k}'} c^{\dagger}_{\mathbf{k}'\uparrow} c^{\dagger}_{-\mathbf{k}'\downarrow} c_{-\mathbf{k}\downarrow} c_{\mathbf{k}\uparrow}, \tag{26.30}$$

where $V_{\mathbf{k},\mathbf{k}'} < 0$ and $\epsilon_{\mathbf{k}} = 0$ at the Fermi level of the normal metal. In terms of the pair creation operator $b^{\dagger}_{\mathbf{k}} = c^{\dagger}_{\mathbf{k}\uparrow} c^{\dagger}_{-\mathbf{k}\downarrow}$, one can write

$$H_{BCS} = \sum_{\mathbf{k}} \epsilon_{\mathbf{k}} (n_{\mathbf{k}\uparrow} + n_{-\mathbf{k}\downarrow}) + \sum_{\mathbf{k},\mathbf{k}'} V_{\mathbf{k},\mathbf{k}'} b^{\dagger}_{\mathbf{k}'} b_{\mathbf{k}}. \tag{26.31}$$

---

[10] Remarkably, Δ has an essential singularity for $V = 0$ and cannot be obtained by a perturbation expansion in powers of V.

[11] Physical Review. 106 (1): 162–164.

[12] Incidentally, Bardeen had already received a Nobel prize with William Schockley and Walter Brattain for the invention (by the use of paper and pencil) of the transistor.

Schrieffer told me during a dinner that he started from the idea of a variational solution for the ground state of the form

$$\Psi_0 \sim \prod_k N_k \exp(g_k b_k^\dagger)|0\rangle, \tag{26.32}$$

where $|0\rangle$is the vacuum, while the real function g_k and the normalization N_k are new unknowns. When he told this to John Bardeen, he exclaimed: "that's the answer!".

This ansatz was suggested to Schrieffer by analogy. Insofar as pairs are analogous to Bosons, this is a coherent state of pairs (see Sect. 16.6). This is essentially correct, but pairs are not true Bosons, since $(b_k^\dagger)^2 = 0$. Therefore, the Fermionic coherent state is obtained from the beginning of the series expansion of the exponential in (26.32), that is, is taken to be of the form:

$$\Psi_0 = \prod_k N_k(1 + g_k c_{\mathbf{k}\uparrow}^\dagger c_{-\mathbf{k}\downarrow}^\dagger)|0\rangle, \tag{26.33}$$

with the normalization condition

$$N_{\mathbf{k}} = \frac{1}{\sqrt{1 + g_{\mathbf{k}}^2}}. \tag{26.34}$$

But the solution (26.33) does not have the same symmetry as the problem. Indeed, the Hamiltonian (26.30) has U(1) symmetry, that is, it is invariant when one multiplies the spin-orbital $\overrightarrow{k}\,\sigma$ by a phase factor of the form $exp(i\theta)$ with real θ; instead, (26.33) is not invariant, so the solution is less symmetric than the problem. This is a famous example of *spontaneously broken symmetry* as discussed above. Historically, this example was important for the development of the theory of the Higgs Boson.

Another remarkable feature of this solution is that it does not have a sharp number of electrons (in analogy with the Boson case presented in Chap. 16). Letting $u_k = N_k$, $v_k = g_k N_k$, Eq. (26.33) becomes

$$\Psi_0 = \prod_k (u_{\mathbf{k}} + v_{\mathbf{k}} c_{\mathbf{k}\uparrow}^\dagger c_{-\mathbf{k}\downarrow}^\dagger)|0\rangle. \tag{26.35}$$

So, $v_{\mathbf{k}}$ is the amplitude that the pair is present and $u_{\mathbf{k}}$ is the amplitude that the pair $c_{\mathbf{k}\uparrow}^\dagger c_{-\mathbf{k}\downarrow}^\dagger|0\rangle$ is absent in the ground state. The pairing is total (the two states are either both empty or both filled). In the normal state (Fermi liquid) $u_{\mathbf{k}} v_{\mathbf{k}} \equiv 0$. Instead, the occupation probability of electron states in the superconductor is given by

$$\langle\Psi_0|\hat{n}_{\mathbf{p}\uparrow}|\Psi_0\rangle = \langle 0|(u_p + v_p b^\dagger) n_{\mathbf{p}\uparrow}(u_p + v_p b^\dagger)|0\rangle = v_{\mathbf{p}}^2, \tag{26.36}$$

while it is a simple exercise to show that

$$\langle\Psi_0|b_{\mathbf{p}}^\dagger b_{\mathbf{q}}|\Psi_0\rangle = u_{\mathbf{p}} v_{\mathbf{p}} u_{\mathbf{q}} u_{\mathbf{q}}. \tag{26.37}$$

Let us solve the Schrödinger equation variationally, by minimizing

$$\langle\Psi_0|H_{BCS}-\mu\sum_{\mathbf{k}}(n_{\mathbf{k}\uparrow}+n_{-\mathbf{k}\downarrow})|\Psi_0\rangle=\sum_k 2(\epsilon_k-\mu)v_k^2+\sum_{pq}V_{pq}u_p v_p u_q v_q \tag{26.38}$$

where the Lagrange multiplier μ fixes the number of electrons. Setting $v_k = \sin(\phi_k)$, $u_k = \cos(\phi_k)$, we find

$$E-\mu N=\sum_k 2(\epsilon_k-\mu)\sin(\phi_k)^2+\frac{1}{4}\sum_{p,q}V_{p,q}\sin(2\phi_p)\sin(2\phi_q), \tag{26.39}$$

and the condition of stationarity $\frac{\partial(E-\mu N)}{\partial\phi_k}=0$ becomes

$$4(\epsilon_k-\mu)\sin(\phi_k)\cos(\phi_k)+\frac{1}{4}\left(\sum_q 2V_{pk}\sin(2\phi_p)\cos(2\phi_k))=0\right),$$

that is

$$2(\epsilon_k-\mu)\sin(2\phi_k)+\cos(2\phi_k)\sum_q V_{kq}\sin(2\phi_q)=0.$$

Then one defines the gap parameter:

$$\Delta_p=-\sum_q V_{pq}u_q v_q=-\frac{1}{2}\sum_q V_{pq}\sin(2\phi_q). \tag{26.40}$$

The minimum condition reads $(\epsilon_k-\mu)\sin(2\phi_k)=\cos(2\phi_k)\Delta_k$ that is

$$\tan(2\phi_k)=\frac{\Delta_k}{\epsilon_k-\mu.} \tag{26.41}$$

We can eliminate ϕ_k using the identity

$$2u_k v_k=\sin(2\phi_k)=\frac{\tan(2\phi_k)}{\sqrt{1+\tan^2(2\phi_k)}}=\frac{\Delta_k}{\sqrt{(\epsilon_k-\mu)^2+\Delta_k^2}}\equiv\frac{\Delta_k}{E_k}, \tag{26.42}$$

where

$$E_k=\sqrt{\epsilon_k-\mu)^2+\Delta_k^2}. \tag{26.43}$$

Substituting $\sin(2\phi_k) = \frac{\Delta_k}{E_k}$ in Eq. (26.40), one obtains the gap equation

$$\Delta_p = -\sum_q V_{pq} \frac{\Delta_q}{2E_q}. \tag{26.44}$$

E_k is a modified dispersion law for the Fermi particles; in the absence of pairing $E_k = 0$ at the Fermi level, but the presence of Δ_k produces a *gap* in the density of states, which modifies the Fermi distribution and can be observed in photoemission experiments. In practice one can take $\mu = 0$, since in the normal $\rightarrow$ superconductor phase transition the correlation energy per particle changes by as little as 10^{-8} eV. This is quite remarkable: a drastic change of the system properties is induced by a quite small energy gain.

26.5.5.1 "New" Superconductors

The above is the mechanism of the so-called conventional superconductors, i.e., metals and alloys that start to superconduct below a critical temperature T_C which is of the order of several degrees Kelvin at most. The high-T_C superconductors discovered in 1986 by Georg Berdnoz and Alex Müller (Nobel prize 1987) reach 138°, a strong progress but not enough for many applications. The mechanism in this case could be different (among others, the writer has proposed one, based on symmetry considerations), but there is a strong disagreement and (alas) some animosity in the community. But much else is worth investigating. There are also anomalous superconductors like UTe_2 that have spin triplet pairing. There are also Heavy Fermion superconductors.[13] In summary, there is still controversy about the mechanism leading to the attraction in different cases, but once an attraction mechanism is given, the general pattern of the BCS theory works.

26.5.5.2 Simplified BCS Hamiltonian

Schrieffer[14] proposed the simplified model in which the BCS interaction V_{pq} vanishes except for both ϵ_p and ϵ_q within a belt wide $2\omega_D$ around the Fermi level, where

$$V_{pq} = -V < 0. \tag{26.45}$$

Then, within the belt, $\Delta_p > 0$; substituting in the gap equation and changing the sum into an integral,

$$\Delta_0 = V \int_{-\omega_D}^{\omega_D} d\epsilon \rho(\epsilon) \frac{\Delta_o}{2E_q} \rightarrow 1 = V \int_{-\omega_D}^{\omega_D} d\epsilon \rho(\epsilon) \frac{1}{2\sqrt{\epsilon^2 + \Delta_0^2}}. \tag{26.46}$$

[13] Aoki et al. (2012).

[14] In "Theory of Superconductivity", Frontiers in Physics, Addison-Wesley (1964).

Assuming a constant ρ, we obtain $\frac{1}{\rho V} = \int_0^{\omega_D} d\epsilon \frac{1}{\sqrt{\epsilon^2+\Delta_0^2}}$ and since

$$\int \frac{dx}{\sqrt{x^2+\Delta^2}} = ln(x+\sqrt{x^2+\Delta^2}) + constant \tag{26.47}$$

$$\frac{1}{V\rho} = \ln[\frac{\omega_D}{\Delta_0} + \sqrt{1+\frac{\omega_d^2}{\Delta_0^2}}]. \tag{26.48}$$

This is solved by

$$\Delta_0 = \frac{\omega_D}{\sinh(\frac{1}{\rho V})}. \tag{26.49}$$

For small ρV (weak coupling), this formula resembles Cooper's binding energy.

Substituting the solution in (26.39) with $\mu = 0$ one obtains the energy of the superconducting state, to be compared to the normal state $E_N = sum_{\epsilon_k < E_F} 2\epsilon_k$. The condensation energy at weak coupling turns out to be

$$C = E_N - E = \frac{1}{2}\rho\Delta_0^2. \tag{26.50}$$

The critical field H_C is such that

$$\frac{H_C^2}{8\pi} = C, \tag{26.51}$$

that is, its energy density is equal to the condensation energy.

26.5.5.3 Bogoliubov–Valatin Transformation

Given that the totally paired state (after Eq. 26.35)

$$\Psi_0 = \prod_k (u_k + v_k c^\dagger_{\vec{k}} c^\dagger_{-\vec{k}})|0\rangle \tag{26.52}$$

represents the ground state of the BCS superconductor model, which Fermion operators create its excitations? We must seek operators that annihilate the ground state, and use a diabolical trick, letting electrons appear out of nothing and disappear. This is a simple way to let our superconductor exchange electrons with (the Fermi level $E_F = 0$ of) a normal metal. Now, notice that

$$u_k c_{k\uparrow}(u_k + v_k c^\dagger_{k\uparrow} c^\dagger_{-k\downarrow})|0\rangle = u_k v_k c^\dagger_{-k\downarrow}|0\rangle$$

$$v_k c_{-k\downarrow}(u_k + v_k c^\dagger_{k\uparrow} c^\dagger_{-k\downarrow})|0\rangle = u_k v_k c^\dagger_{-k\downarrow}|0\rangle.$$

So,

$$(u_k c_{k\uparrow} - v_k c_{-k\downarrow})(u_k + v_k c^{\dagger}_{k\uparrow} c^{\dagger}_{-k\downarrow})|0\rangle = 0. \tag{26.53}$$

We conclude that the Fermion operator

$$\alpha_k = u_k c_{k\uparrow} - v_k c^{\dagger}_{-k\downarrow} \tag{26.54}$$

has the ground state Ψ_0 as a vacuum. Thus, α_k destroys excitations and changes the number of electrons. There is another one, namely the orthogonal combination

$$\beta_k = u_k c_{-k\downarrow} + v_k c^{\dagger}_{k\uparrow}. \tag{26.55}$$

Hence, the excitations, called Bogolons, are created by

$$\alpha^{\dagger}_k = u_k c^{\dagger}_{k\uparrow} - v_k c_{-k\downarrow} \tag{26.56}$$

$$\beta^{\dagger}_k = u_k c^{\dagger}_{-k\downarrow} + v_k c_{k\uparrow}. \tag{26.57}$$

This is the Bogoliubov–Valatin transformation. One can readily verify that it is canonical ($[\alpha_k, \alpha^{\dagger}_{k'}]_- = \delta(k, k') = [\beta_k, \beta^{\dagger\prime}_k]_-$). The inverse transformation is

$$c_{k\uparrow} = u_k \alpha_k + v_k \beta^{\dagger}_k, \qquad c_{-k,\downarrow} = u_k \beta_k - v_k \alpha^{\dagger}_k$$

$$c^{\dagger}_{k\uparrow} = u_k \alpha^{\dagger}{}_k + v_k \beta_k, \qquad c^{\dagger}_{-k\downarrow} = u_k \beta^{\dagger}{}_k - v_k \alpha_k.$$

Acting with the annihilation operator of any excitation on the ground state Ψ_0 yields 0.

26.5.5.4 Excited Pairs

It is a simple exercise to show that the pairs have an excited state. One can excite a pair, conserving the number of electrons:

$$\phi^{exc}_k = \alpha^{\dagger}_k \beta^{\dagger}_k (u_k + v_k c^{\dagger}_{k\uparrow} c^{\dagger}_{-k\downarrow})|0\rangle = (u_k c^{\dagger}_{k\uparrow} c^{\dagger}_{-k\downarrow} - v_k)|0\rangle. \tag{26.58}$$

26.5.6 Why Perfect Diamagnetism and Zero Resistance?

It is quite common that students go through the intricacies of the BCS theory but fail to gain a simple intuitive understanding of the basic properties (zero resistance, London rigidity, perfect diamagnetism) in terms of Cooper pairs. But a formalism without physical understanding is not rewarding. Here is a simple qualitative explanation.

In a normal metal, where the Fermi distribution holds, every Fourier component of the field carrying wave vector $\overrightarrow{q}$ couples the electron wave functions having wave vectors $\overrightarrow{k}$ and $\overrightarrow{k} + \overrightarrow{q}$. Such random perturbations are introduced by the quantized lattice vibrations (phonons) and, together with impurities and other imperfections, lead to energy dissipation and resistance. Consider the current

$$\boldsymbol{J}_0(\boldsymbol{k}) = \frac{e\hbar}{2mi}(\phi_{\boldsymbol{k}}^{\dagger}\nabla\phi_{\boldsymbol{k}} - \phi_{\boldsymbol{k}}\nabla\phi_{\boldsymbol{k}}^{\dagger})$$

calculated over a plane-wave spinor $\phi_{\boldsymbol{k}} = \frac{\exp(i\boldsymbol{k}\boldsymbol{x})}{\sqrt{\Omega}}\chi$, where Ω is the volume and χ a spin function. It is $\boldsymbol{J}_0(\boldsymbol{x}) = \frac{e\boldsymbol{k}}{m\Omega}$; this is a vector, and its average over the ground state of a superconductor that has no privileged direction vanishes.

Let us switch an external field with $H' = -\frac{e}{mc}\overrightarrow{A}.\overrightarrow{p}$, in the gauge $\operatorname{div}\overrightarrow{A} = 0$. The average of the momentum operator $\overrightarrow{p}$ over a pair state $(\boldsymbol{k}, -\boldsymbol{k})$ is obviously zero; we know (Sect. 19.3) that $\overrightarrow{p}$, being a one-body operator, has vanishing matrix elements between pairs with different $\boldsymbol{k}$. In other terms, it cannot take (in first-order) from a pair state $c_{k\uparrow}^{\dagger}c_{k\downarrow}^{\dagger}|0\rangle$ to a different one $c_{k'\uparrow}^{\dagger}c_{k'\downarrow}^{\dagger}|0\rangle$ with $k \to k'$. Only the excited pair ϕ_k^{exc} remains to be considered, but even in this case it is evident that no first-order contribution arises from H'.

This removes the main mechanism producing resistance[15] and explains the London rigidity of the wave function (this expression means that the ground state is not altered by any weak perturbation).

So, there are no matrix elements between the ground state and the low (paired) excited states; one has to break pairs in order to excite the system, but this costs an energy exceeding the gap. So, there is no first-order correction to the wave function, and the ground state is not changed by the application of a weak external field. The wave function of the superconductor is rigid; and from this very rigidity, the characteristic properties of superconductors descend. It is known as London Rigidity, after a theory developed by brothers Fritz and Heinz London. The electron current density in the presence of the field is obtained by adopting the minimal coupling (Chap. 2):

$$\boldsymbol{J}(\boldsymbol{x}) = \boldsymbol{J}_0(\boldsymbol{x}) - \frac{e^2}{mc}\Psi^{\dagger}\boldsymbol{A}\Psi; \tag{26.59}$$

the extra term is diamagnetic. In a normal metal, the carriers are electrons, and a magnetic field couples to their spin; so, the first-order correction to the many-body wave function is responsible for the fact that in the presence of an external field, a paramagnetic current flows and prevails over the diamagnetic contribution. In superconductors,[16] however, the carriers are spinless pairs, where $\mathbf{k}$ and $-\mathbf{k}$ enter symmetrically and therefore $\langle \boldsymbol{J}_0(\boldsymbol{x})\rangle = 0$ even in the presence of the field; therefore

[15] Actually in superconductors the resistance is zero, or at least, the experimental results are consistent with zero.

[16] I leave aside the rare, more involved case of triplet superconductors, like Sr_2RuO_4.

$$\boldsymbol{J}(\boldsymbol{x}) = -\frac{ne^2}{mc}\boldsymbol{A}(\boldsymbol{x}). \tag{26.60}$$

Equation (26.60) is just the London equation.[17] As we have seen above, it leads to the Meissner effect and describes the phenomenology of the superconductor, namely zero resistance and perfect diamagnetism. A curl gives us

$$rot\,\boldsymbol{J} = -\frac{ne^2}{mc}\boldsymbol{B}; \tag{26.61}$$

a second curl (using rot rot = grad div $-\nabla^2$) leads to

$$\nabla^2\boldsymbol{B} = \frac{4\pi ne^2}{mc^2}\boldsymbol{B}. \tag{26.62}$$

This yields the perfect diamagnetism; indeed, solving for a half space of a superconductor with a magnetic field applied on the vacuum side, the field is found to decay exponentially inside the superconductor. The London penetration depth is $\Lambda = \sqrt{\frac{mc^2}{4\pi ne^2}}$. A time derivative of (26.61) then yields

$$rot\,\frac{\partial \boldsymbol{J}}{\partial t} = -\frac{ne^2}{mc}\frac{\partial \boldsymbol{B}}{\partial t}. \tag{26.63}$$

Since $\boldsymbol{J} = ne\boldsymbol{v}$ and $\frac{\partial \boldsymbol{B}}{\partial t} = c\nabla \times \boldsymbol{E}$, this implies that $m\frac{\partial \boldsymbol{v}}{\partial t} = e\boldsymbol{E}$, the electric field accelerates the charges, and there is no damping. The resistance vanishes.

The above is just a pedagogical account of the main physical ideas involved in the BCS theory. The subject is very rich. A magnetic field exceeding a critical value H_c kills superconductivity; in type I superconductors, the metal returns normal, while in type II superconductors there is a mixed situation described by the theory of Ginzburg, Landau, Abrikosov and Gorkov. In addition, there are (relatively) high-temperature ceramic superconductors with several theoretical approaches (including one by Michele Cini and Adalberto Balzarotti[18]) and much ado for searching room temperature superconductors.

[17] The London equation was proposed by brothers Fritz and Heinz London in 1935.

[18] Cini and Balzarotti (1997).

26.5.7 The Phenomenological Richardson Model for Superconductivity

Any microscopic theory is necessarily involved, but a much simpler and phenological description credited to Richardson[19] has the merit of being readily applicable to finite systems like nanoscopic clusters and even atomic nuclei, where a similar pairing between nucleons exists. The Hamiltonian is taken to be

$$H = \sum_{k,\sigma} \epsilon_{\boldsymbol{k}} \hat{n}_{\boldsymbol{k},\sigma} + g \sum_{\boldsymbol{k},\boldsymbol{h}} c^{\dagger}_{\boldsymbol{k}+} c^{\dagger}_{-\boldsymbol{k}-} c_{-\boldsymbol{h}-} c_{\boldsymbol{h}+}, \tag{26.64}$$

where the first term is kinetic energy, while the interaction term scatters a pair of Fermions of opposite spin and wave vector $\overrightarrow{h}$ to a similar pair with a different wave vector. This model Hamiltonian is tantamount to an assumption that some pairing mechanism is in action. The unpaired electrons remain spectators. Therefore, we forget about them; the dynamics involves pairs, and we introduce pair operators

$$b^{\dagger}_{\boldsymbol{k}} = c^{\dagger}_{\boldsymbol{k},\sigma} c^{\dagger}_{-\boldsymbol{k},-\sigma}. \tag{26.65}$$

We rewrite the remaining pair model as

$$H_{pair} = \sum_{kh} (2\epsilon_h \delta_{hk} - g) b^{\dagger}_h b_k. \tag{26.66}$$

The pair operators have commutator relations

$$b^{\dagger 2}_j = 0, \tag{26.67}$$

$$[b_h, b^{\dagger}_k] = \delta_{hk}(1 - 2b^{\dagger}_k b_k), \tag{26.68}$$

$$[b^{\dagger}_k b_k, b^{\dagger}_h] = \delta_{hk} b^{\dagger}_k. \tag{26.69}$$

These are called hard-core boson relations. Physically, one can consider that these relations are Boson relations in almost all cases, and it is a good idea to treat the b operators as if they were hundred percent Bosons, with $[b_k, b^{\dagger}_h] = \delta_{hk}$ instead of being too choosy. The model Hamiltonian then becomes the solved model of Sect. 22.5. The Hamiltonian is diagonalized in the form

$$\tilde{H} = \sum_k E_k B^{\dagger}_k B_k \tag{26.70}$$

[19] See Richardson (1963); also Jan Von Delft and Fabian Braun, cond-mat/9911058.

and the interacting ground state is just the ground state of a system of renormalized bosons $B_k^\dagger$; one finds that

$$B_k^\dagger = gC_k \sum_h \frac{b_h^\dagger}{2\epsilon_k - E_h} \tag{26.71}$$

and

$$\frac{1}{(gC_k)^2} = \sum_h \frac{1}{2\epsilon_h - E_k}. \tag{26.72}$$

So, the ground state is made up of pairs and the ground state energy for n pairs is the sum of n eigenvalues found in Sect. 22.5. This is negative, so the Fermi distribution is superseded by the paired state. Incidentally, the Richardson model with hard-core bosons is also solved exactly and does not modify the qualitative picture.

26.6 The Integer Quantum Hall Effect and Topological Insulators

The Hall effect discussed in Chap. 4 had metal samples and originally aimed at the effects of the Lorentz force on the (then unknown) charge carriers. The field was found to produce a transverse resistance linear in the magnetic field; it was important fundamental research in 1879, well before the discovery of the electron. By Quantum Hall effects one does not refer to the quantum theory of the effect discovered by Hall, but to phenomena which were discovered in 1980 and later, still involving a current in a very strong magnetic field in thin film materials. Roughly speaking, at very low temperatures and strong magnetic fields, in some materials the electrons get bound to the magnetic flux quanta in various ways, which are detected by the Hall measurements.

In 1980 the Integer Quantum Hall effect was discovered experimentally, not predicted by anybody (what a lost occasion for us theoreticians!). It opened the door to unexpected developments, and it also has considerable importance in fundamental research, allowing, among others, a precise measurement of Planck's constant, as mentioned at the beginning of this chapter. It differs from the classical effect because of the quantum nature of the phenomena and the experimental conditions, and was allowed by technological advances (in a way, recalling the discovery of superconductivity by Onnes, who also was the first to reach the temperatures needed for the phenomenon to occur). In the original experiment by Klaus von Klitzing and coworkers, the sample was a very thin Si junction, the temperature was about one Kelvin, and the field $\vec{B}$ very strong (larger than 10 T). The Hall resistivity was not linear in the magnetic field as expected classically, but showed a sharp staircase growth due to quantization.

26.6.1 Landau Levels in a Quasi Two-Dimensional System

Let us represent our sample for the Hall measurement as a thin film lying on the xy plane, with the current along x, the magnetic field parallel to the z axis and the Hall field along the y axis. As a preliminary calculation, we need the free-electron density of states of the thin film, neglecting spin (for the moment); let us also neglect the electron–electron interaction[20] and any potential. The total one-particle density of states (different from the local one of Chap. 22) is defined as:

$$\rho(E) = \sum_{\vec{k}} \delta(E - \frac{\hbar^2 \vec{k}^2}{2m}) \tag{26.73}$$

where $\vec{k}$ is the two-dimensional wave vector; we do not care about the z direction because the film is so thin that the excitations involving it have high energy and can be neglected for our purposes. For a rectangular sample with sides L_x and L_y, one readily finds:

$$\rho(E) = \frac{2\pi m S}{h^2}\theta(E), \tag{26.74}$$

where $S = L_x L_y$. Thus the density of states (DOS) $D(E)$ is a constant for positive E.

Next we introduce the magnetic field $\vec{B}$ parallel to the z axis. We find Landau levels like in Paragraph 17.2 except that there is no propagation along the z axis.

The density of states in the presence of a magnetic field is a series of delta functions (see Fig. 26.4). The magnetic field is always a strong perturbation, since it changes a continuous spectrum to a discrete spectrum.

The density of states is of the form:

$$\rho(E) = D_L(B)\Theta(E)\sum_{n>0}\delta(E - (n + \frac{1}{2})\hbar\omega_C), \tag{26.75}$$

where $D_L(B)$ is the degeneracy of each Landau level.

Calculation of the degeneracy

$D_L(B)$ can be obtained by counting the number of degenerate solutions. In the Landau gauge $\vec{A} = B(-y, 0, 0)$, with periodic boundary conditions along x, we use the model Hamiltonian:

$$\hat{H} = \frac{1}{2m}\left[\left(p_x - \frac{eBy}{c}\right)^2 + p_y^2\right] + eE_0 y, \tag{26.76}$$

[20] We are treating the electrons as quasi-particles in the spirit of Sect. 19.4. As discussed there, this approach is successful in many cases but fails in strongly correlated systems. It also fails in the case of the fractional quantum Hall effect (see below).

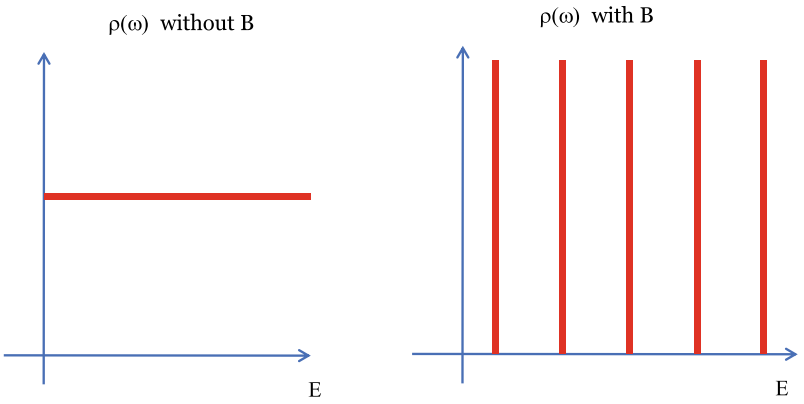


Fig. 26.4 Density of states of the two-dimensional non-interacting electron gas. Left: no field. Right: qualitative effects of the magnetic field

where B is the magnetic field, while the transverse electric field E_0 is the Hall field; however, we shall drop E_0 at the end of the calculation since we wish to characterize the unperturbed system. Since p_x is conserved, the wave functions are plane waves along x, with $p_x = \hbar k_x$ and $k_x = \frac{2\pi n}{L_x}$ with n integer.

Now, we try a solution of the form

$$\psi_{kn} = e^{ikx}\phi_n(y - y_0), \tag{26.77}$$

where ϕ_n is a harmonic oscillator wave function. Expanding,

$$H = \frac{p_y^2}{2m} + \frac{1}{2}m\omega_c^2 y^2 + \frac{k_x^2}{2m} + (-\hbar k_x \omega_c + 2E_0)y, \tag{26.78}$$

where

$$\omega_c = \frac{eB}{mc}. \tag{26.79}$$

Next, we find the shift y_0 such that

$$\frac{1}{2}m\omega^2{}_c y^2 + (-\hbar k_x \omega_c + eE_0)y = \frac{1}{2}m\omega_c^2(y - y_0)^2 - y_0)^2 - \frac{1}{2}m\omega_c^2 y_0^2.$$

In this way we obtain a shifted oscillator + a constant. Expanding the right-hand side, one gets an equation for y_0 which is solved by

$$y_0 = \frac{1}{\omega_c}\left(\frac{\hbar k_x}{m} - c\frac{E_0}{B}\right). \tag{26.80}$$

Without the second term in E_0, which is the Hall contribution proportional to the current, this result is related to a fundamental quantum property since

$$y_0 = \frac{hc}{e}\frac{k_x}{2\pi B} = \phi_0 \frac{k_x}{2\pi B}, \tag{26.81}$$

where $\phi_0 = \frac{hc}{e} = 4.136 \times 10^{(-7)}$ Gauss cm^2 is the magnetic flux quantum that we have met already in Sect. 22.3. So the solutions are plane waves along the x axis but oscillators along y, and the minimum along y, $y_0 = \frac{c\hbar k_x}{eB}$, is proportional to k_x. But for the n-th state of the s-th LL, $k_x = \frac{2\pi n}{L_x}$ and $y_0 = \frac{c\hbar k_x}{eB} = \frac{c\hbar \frac{2\pi n}{L_x}}{eB} \leq L_y$. It turns out that

$$n \leq \frac{eB}{chL_xL_y} \tag{26.82}$$

so setting the sample area $S = L_xL_y$ and recalling the fluxon $\phi_0 = \frac{hc}{e}$, the degeneracy is

$$D_L = \frac{eBS}{hc} = \frac{\phi}{\phi_0}. \tag{26.83}$$

Electrons pick up fluxons

This implies that if the number of electrons is equal to the degeneracy D_L, which amounts to putting all the particles in the ground state, each particle picks up a fluxon. In other terms, the condition is $H_tS = N_{el}\phi_0$. Typically, for 10^{11} electrons per square centimeter, this is achieved with a threshold field $H = H_t = 10^4$ Gauss.

But we *forgot* about spin. Introducing spin, we must fulfill the condition for both spins, so we require that

$$B_tS = \frac{N_{el}}{2}\phi_0. \tag{26.84}$$

Decreasing the field below B_t, one starts filling the LL (Landau levels) with n = 1. At $B = \frac{B_t}{2}$ the second LL n = 1 is full. The LL number $n = \nu$ is partially filled between $B = \frac{B_t}{\nu}$ and $B = \frac{B_t}{\nu+1}$. The Landau levels $E_n = (n + \frac{1}{2}\hbar\omega_c)$ grow linearly with H since $\omega_c = \frac{eB}{mc}$.

26.6.2 The Staircase Paradox and the Revenge of the Surface

The Hall resistance ρ_{xy} is proportional to B in the Drude theory, and this is expected classically because the Lorentz force proportional to B acts on the charge carriers and produces the shift of the charge distribution which is the source of the electric field along y. Actually this is about right, but the plot of resistance versus B shows a spectacular staircase trend superimposed on the linear trend. A sketch is shown in the next figure.

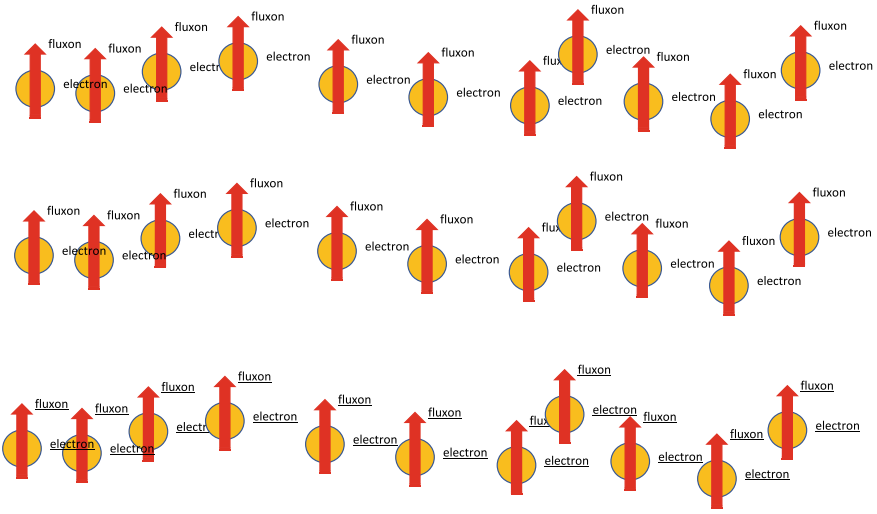


Fig. 26.5 When all electrons fill exactly the lowest Landau level (LLL), each electron gets a fluxon

Let ν denote the number of filled LL, which is a function of the field B. For integer ν, there are only filled or empty LL, the Fermi level is in a gap, and by analogy with the band theory of solids (Sect. 17.7) one should expect a band insulator. But the experiment shows that, quite the other way, the xx resistivity is 0. The whole experimental situation looks paradoxical.

For $\nu = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$ we have a half-filled LL and one should expect a metal, but experiment shows maximum xx resistivity.

In the intervals between $B = \frac{B_t}{\nu}$ and $B = \frac{B_t}{\nu+1}$. where a LL is partially filled, the longitudinal (xx) resistivity is very low and when there is an integer number of filled LL, the Hall (xy) resistivity is constant with an astonishing precision, such that it allowed to improve the measurement of Planck's constant (Figs. 26.6 and 26.7).

But in Sect. 17.7 we had an implicit assumption that seemed obvious, an assumption that works in countless cases, and that is that the surface doesn't matter. The resolution of the paradox is that the measured current is carried essentially by the surfaces and the plateaux have a topological nature (see below).

Very remarkably, one gets the correct step height from Drude theory, assuming an integer number of occupied LL and using the fact that the longitudinal resistance vanishes at the steps. The Hall resistance in Drude theory is, in cgs units:

$$\rho(B) = \begin{pmatrix} \rho_{xx} & \rho_{xy} \\ \rho_{yx} & \rho_{yy} \end{pmatrix} = \frac{B}{|e| n_s c} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{26.85}$$

where n_s is the electron number density appropriate for the surface. So,

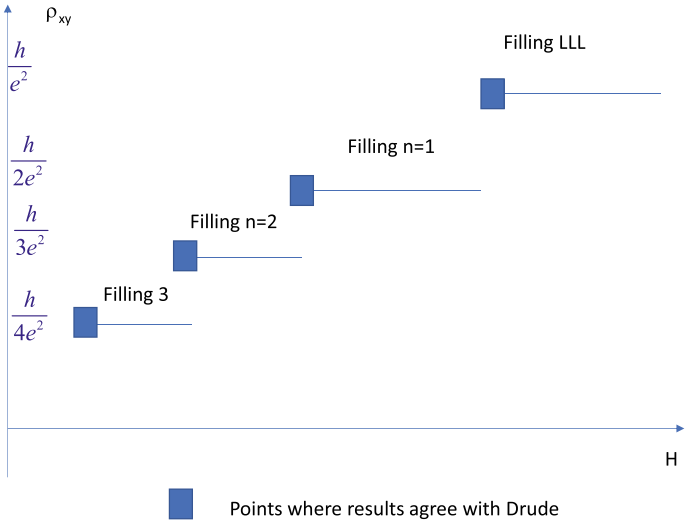


Fig. 26.6 Qualitative magnetic field dependence of the Hall resistance in the integer Quantum Hall Effect regime. Right at the steps (blue squares), the xx resistivity vanishes. The field needed to put all electrons in the LLL is of order of tens of kilogauss

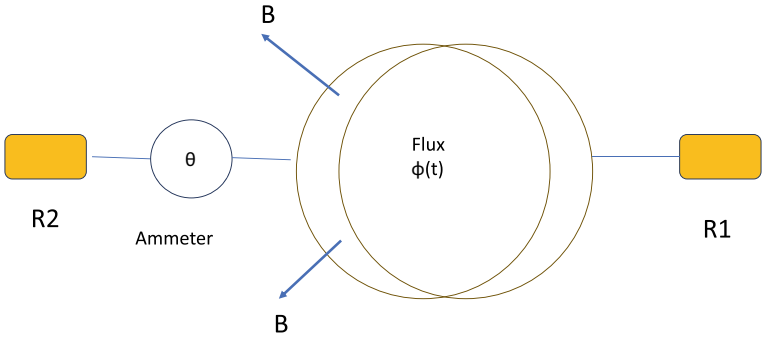


Fig. 26.7 The ring pierced by the field $\overrightarrow{B}$ and the flux ϕ is connected to the reservoirs A and B, and in the connection to A an ammeter measures the current; the reading is given by the angle θ

$$\rho_{xy} = \frac{B}{en_s c}, \tag{26.86}$$

where n_s is the surface density of electrons. Now, we must evaluate the field B at the steps.

The condition for having all the $N_{el} = N_\uparrow + N_\downarrow$ electrons of both spins in the LLL is

$N_\uparrow = N_\downarrow = D_L$ where $D_L = \frac{\phi}{\phi_0}$ is the number of fluxons through the sample.

$H_t S = \frac{1}{2\phi_0} N_{el}$. Since all the LL have the same degeneracy, ν levels are exactly filled with $B = \frac{B_t}{\nu}$. This is in good agreement with several features of the experimental pattern, but why? And why the plateaux? Why so incredibly exact?

26.6.3 Topological Explanation of the Sharp Plateaux

The explanation was first given in a remarkable paper by Thouless and coworkers[21] and is highly innovative and very interesting. Below, I illustrate a simpler variant by Avron and coworkers.[22] The specimen is folded to form a ring inside which a time-dependent flux $\phi(t)$ produces a tangential electric field proportional to $\dot{\phi}$ (by the Faraday effect) and excites the current. This field acts as E_x in the standard geometry of Sect. 4.5.2, where we met the classical Hall effect. The ring is also subjected at each point to a second radial magnetic field perpendicular to the ring itself. This acts as B_z in Sect. 4.5.2 and produces a Hall electric field which is due to the fact that the Lorentz force acting on the charge carriers produces a transverse polarization of the sample.[23] In Sect. 4.5.2 this field is E_y. Each of the two sides of the sample is connected to a separate infinite charge reservoir (see Fig. 26.5).[24] Between one of the sides and the reservoir an amperometer measures the Hall current from the sample to the reservoir, and its reading is given by an angle θ. This might represent the effect of a measurement of the Hall potential with an ammeter.

The Hamiltonian is taken to be

$$H = H(\phi, \theta) \tag{26.87}$$

where ϕ is the flux in the ring and θ is the reading in the ammeter that measures the Hall current J. Thus, (ϕ, θ) constitute the parameter space. The x axis is replaced by the curved coordinate around the ring, and we shall denote it as J_x. The current in the ring J_x and also the Hall current J are proportional to $\frac{\partial \phi}{\partial t}$. Now we introduce the Hamiltonian $H = H(\theta, \phi)$ of the system and its wave function $\psi(\theta, \phi)$.

Don't worry! We do not have to solve the Schrödinger equation, which would be hard, but only to find the topological reason why the parameter space curvature explains quantization. Since the experiment is performed slowly, we can use an adiabatic approximation:

---

[21] Thouless et al. (1982); Thouless received the Nobel prize in 2016 jointly with F. Duncan, M. Haldane, and Michael Kosterlitz.

[22] Joseph Avron, Daniel Osadchy, and Ruedi Seiler, Physics Today, August 2003.

[23] This geometry is almost the same as was first proposed by Laughlin (1981); that paper was very useful, but did not explain the quantization completely because it did not use the topological reasoning to be explained below.

[24] Generally the circuit should be closed, but since the reservoirs are supposed to be infinite, this fact is not important here.

$$H\psi = i\hbar\frac{\partial\psi}{\partial t} = i\hbar\frac{\partial\psi}{\partial\phi}\frac{\partial\phi}{\partial t}. \tag{26.88}$$

Now it is clear physically that the angle θ of the ammeter measuring the Hall current does not change the energy. So,

$$\frac{\partial}{\partial\theta}\langle\psi|H|\psi\rangle = 0, \tag{26.89}$$

that is

$$\langle\frac{\partial}{\partial\theta}\psi|H|\psi\rangle + \langle\psi|\frac{\partial}{\partial\theta}H|\psi\rangle + \langle\psi|H|\frac{\partial}{\partial\theta}\psi\rangle = 0, \tag{26.90}$$

and inserting $J = \frac{\partial H}{\partial\theta}$, we find

$$J = -2Re\langle\frac{\partial}{\partial\theta}\psi|H|\psi\rangle = 2Im\langle -i\frac{\partial}{\partial\theta}\psi|H|\psi\rangle. \tag{26.91}$$

Since $H\psi = i\hbar\frac{\partial\psi}{\partial t} = i\hbar\frac{\partial\psi}{\partial\phi}\frac{\partial\phi}{\partial t}$, restoring c,

$$J = \frac{2}{c}Im\langle -i\frac{\partial\psi}{\partial\theta}|i\hbar\frac{\partial\psi}{\partial\phi}\rangle\frac{\partial\phi}{\partial t};$$

so, the Hall current is

$$J = \hbar K\frac{\partial\phi}{\partial t}, \tag{26.92}$$

where

$$K = 2Im\langle\frac{\partial\psi}{\partial\theta}|\frac{\partial\psi}{\partial\phi}\rangle. \tag{26.93}$$

But K is just the curvature of the bundle of ground states. By the Gauss–Bonnet–Chern theorem, K is quantized due the topology of the parameter space (see Sect. 23.3). Hence the current (26.92) that measures the Hall field is also quantized and the Hall resistance is also quantized.

26.6.4 Origin of the Oscillations of Longitudinal Resistivity

Experimentally, the xx resistivity is found to vanish completely at the beginning of each stair step ($n = 1, 2, 3, \ldots$) and to reach maxima at the middle of each step ($n = 0.5, 1.5, 2.5, \ldots$). This is contrary to what is observed normally in solids, where partially filled bands allow metallic conduction, while a Fermi level in a large band gap causes insulating behavior and a small gap causes semiconducting behavior. The accepted explanation for this paradox involves the conduction in edge states.

The presence of the magnetic field imposes a chirality on the motion of conduction electrons; classically they would orbit, all in the same direction. Suppose the electrons can move on the border of a rectangle in a couterclockwise way. Then they will go up on the right side and down in the left side, the motion is ordered, and no head-on collision occurs.This corresponds to zero resistance. At the middle of each step, the charge carriers propagate in the bulk and normal resistance exists.

26.6.5 Hint to Topological Insulators

Striking features of the integer quantum Hall effect are that the samples are insulating in the bulk suggesting an energy gap between valence and conducting states; instead, the edges are conducting; moreover, it is not the electron–electron interaction but the topology that matters. Indeed, several materials behave in a similar way as the samples in the quantum Hall effect without the need for an external magnetic field, and are called topological insulators. A short list includes HgTe, Bi_2Se_3 and $CuInSe_2$. Some are magnetic, like $GdIr_2$ and $MnBi_2$.

Edge states can indeed be very different from the bulk. As an example, we can consider the Su Schrieffer Heeger model that we met in Sect. 17.7.2 and again in Sect. 23.5. For different v and w, this model gives a valence band and a conduction band separated by a gap, as in insulators. However, it turns out that finite chains have zero energy edge states within the gap, with wave functions localized at the ends. These are related to the conducting surface states.

The normal situation in solid state physics is that of an energy band such that the electron wave vector varies in the Brillouin Zone, and the two extremes of the Brillouin Zone are equivalent. In such cases, the topology of the K space is that of a Torus; at the end of the tour, you are again at the beginning. In the traditional texts about Solid State physics, this was a matter of course, taken for granted without the need for explanation.Topological insulators, however, are different because there is a topological hindrance to have equivalent Brillouin Zone boundaries. We met such a situation in a different context in Chap. 15, when discussing the Dirac monopole, and we found that no expression for the vector potential of the monopole can hold for all directions. In that problem, one can conveniently use an expression for the North Hemisphere and a different one for the South Hemisphere. In topological insulators the topology of the Brillouin Zone is such that a reciprocal lattice vector actually takes to a different band, separated from the original band by a gap. This happens, for evident reasons, when the band Hamiltonian is of the form

$$H(k) = h_0(k) + \overrightarrow{h}_1 \cdot \overrightarrow{\sigma}, \tag{26.94}$$

where $\overrightarrow{\sigma}$ stands for the three Pauli matrices; in this case, the situation of Chap. 15 is reproduced. However this happens in the bulk, while at the surface the gap closes

up, so we have a conducting surface and an insulating bulk, like in the Quantum Hall Effect.

26.6.6 Very Short Hint to the Fractional Hall Effect

There is also a fractional quantum Hall effect, for which Robert Laughlin, Horst Störmer and Daniel Tsui received the 1998 Nobel prize. It requires still lower temperatures of the order of 0.1° and magnetic fields of the order of 30 T, with thin semiconductor samples. The Laughlin theory takes into account the electron–electron interaction, which is important in this case. This is outside our scope here; a convenient short account was published by Thomas DeMastri in the Arxiv Cond-Mat in December 2021.

26.6.7 Spin Hall Effect

There is also a spin analogue of the quantum Hall effect, which was predicted in 1971 by Dyakonov and Perel. In some conductors, the current does not produce a Hall field acting on the charges, but polarize spins instead. As a result, opposite boundaries are spin-polarized in opposite directions, or, in the case of cylindrical wires, the spin polarization winds around the surface. The spin Hall effect arises because of the spin–orbit interaction and can be used to manipulate electron spins electrically. Spintronics could be of interest for technology because in principle the energy cost of polarizing spins is much lower than that of moving charges. An article by the discoverer Dyakonov on Cond-mat file:1210.3200.pdf is a useful introduction to this subject. Experimental data on semiconductor heterostructures are reported by Wunderlich et al., on Arxiv: Cond-mat 0410295 (2005).

References

Aoki, D., W. Knafo, I. Sheikin, arXiv:1204.5128v2 (2012)
M. Cini, Phys. Rev. B **89**, 239902(E) (2014)
Cini, M.: Phys. Rev. B **22**, 5887 (1980)
Cini, M., Balzarotti, A.: Phys. Rev. B **56**, 14711 (1997)
Laughlin, R.: Phys. Rev. B **23**, 5632 (1981)
Michele Cini and Enrico Perfetto, PRB 84 245201 (2011)
Richardson, R.W.: Phys. Lett. **3**, 277 (1963)
Thouless, D.: Phys. Rev. B **27**, 6083 (1983)
Thouless, D.J., Kohmoto, M., Nightingale, M.P., den Nijs, M.: Phys. Rev. Lett. **49**, 405 (1982)
van Wees, B.J., et al.: Phys. Rev. Lett. **60**, 848 (1988)
P.W Brower Phys. Rev. B58, 10135 (1998)

Chapter 27
Entanglement, Its Challenges and Its Applications

The theory makes predictions that contrast dramatically with our traditional idea of reality, but all the experimental tests designed to prove the failure of Quantum Mechanics have shown that it works both on microscopic samples and over long distances. It is a reality yet to be explored that promises new surprises and applications.

27.1 What is the Electron Wave Function in the H Atom?

When we dealt with the simplest model for the non-relativistic H atom in Chap. 17, naturally enough we did not start with the electron and the proton; instead, we started with one effective particle having a reduced mass. The elimination of the center of mass degrees of freedom was discussed in the chapters on Classical Mechanics (see Sect. 2.5.1). A similar argument works in Quantum Mechanics, with the momenta replaced with the quantum operators, starting from the two-body Hamiltonian:

$$H = -\frac{\hbar^2}{2m_e}\nabla_e^2 - \frac{\hbar^2}{2m_p}\nabla_p^2 + V(\rho). \tag{27.1}$$

Here, ∇_e acts on r_e, ∇_p acts on r_p, and

$$\overrightarrow{\rho} = \overrightarrow{r_e} - \overrightarrow{r_p}. \tag{27.2}$$

The wave function $\psi = \psi(\overrightarrow{r_e}, \overrightarrow{r_p})$ depends on the coordinates and spins of both particles; however, we omit spins, since H does not depend on them in the non-relativistic limit. Next, we introduce the center of mass

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1_27

$$\overrightarrow{R} = \frac{m_e \overrightarrow{r_e} + m_p \overrightarrow{r_p}}{m_e + m_p}, \tag{27.3}$$

the reduced mass μ with

$$\frac{1}{\mu} = \frac{1}{m_e} + \frac{1}{m_p} \tag{27.4}$$

and the total mass $M = m_e + m_p$. For hydrogen,[1] $\mu \sim m_e$. Using the transformed Hamiltonian

$$H = -\frac{\hbar^2}{2M}\nabla_R^2 - \frac{\hbar^2}{2\mu}\nabla_\rho^2 + V(\rho), \tag{27.5}$$

the problem is *separable* in the sense that

$$\psi(\overrightarrow{r_e}, \overrightarrow{r_p}) = \phi(\overrightarrow{R})\psi(\overrightarrow{\rho}); \tag{27.6}$$

the first factor represents the free motion of the center of mass. Once the quantum mechanical problem is solved, one would be interested in the motions of the electron and of the proton. Classically one can solve Eqs. (27.2) and (27.3) for $\overrightarrow{r_e}$ and $\overrightarrow{r_p}$ and find the trajectories of both particles. One can do so for the orbits of planets and satellites. Since in the quantum problem there are no orbits, one would rather ask for their wave functions.

But here Quantum Mechanics becomes weird. The trouble is that $\psi(\overrightarrow{r_e}, \overrightarrow{r_p})$ is not *separable* in the form $\phi(\overrightarrow{r_e})\psi(\overrightarrow{r_p})$, that is, cannot be written in terms of an electron wave function times a nuclear wave function; moreover, no other way to separate the wave functions of the two particles is allowed. The particles are *entangled*, and such individual wave functions do not exist! For the same reason, in principle, one cannot speak of electron states in molecules and solids separately from the vibrations of the nuclei. In practice one often factors the problem into an electronic problem and a vibrational problem in the *Born–Oppenheimer approximation*, but this approximation may be useful or poor, depending on the problem.

This situation is really very amazing. In everyday life, in order to describe the state of a system, one has to specify the state and the position of the components. In the motor of a car, every screw has a location, and the mechanic checks that everything is in place. In Quantum Mechanics, a complete knowledge of the state of the system generally leaves the subsystems undetermined. When the state of the system is completely known, in general one can only speak of probability of the parts being in a given state.

[1] In Positronium, the proton is replaced by a positron, the antiparticle of the electron, having the same mass, and the reduced mass is $\mu = \frac{m_e}{2}$. It decays into photons.

27.2 EPR Paradox

Another amazing example of entanglement is enlightening. Consider a spin $\frac{1}{2}$ and the eigenfunctions α with eigenvalue $S_z = \frac{1}{2}\hbar$ and β with $S_z = \frac{-1}{2}\hbar$. The two-body basis contains the four states

$$\alpha(1)\alpha(2), \alpha(1)\beta(2), \beta(1)\alpha(2), \beta(1)\beta(2).$$

Now, the state $\alpha(1)\beta(2)$ has particle 1 in α and particle 2 in β, but S^2 has no sharp value. On the other hand, the singlet

$$|\Psi\rangle \equiv |0,0\rangle = \frac{1}{\sqrt{2}}(|\alpha(1)\beta(2)\rangle - |\beta(1)\alpha(2)\rangle) \tag{27.7}$$

which has zero total spin, is a well specified two-body state, but is entangled: none of the particles have a precise state. However, if the pair is prepared as $|0,0\rangle$ and then the spin component of one particle is measured, the spin component of the second particle is also determined. This might appear as a familiar situation, comparable to the trivial fact that if a glove is left, its companion must be right. However, the situation is very different and paradoxical. Suppose the system is prepared in the state $|S, M\rangle = |0,0\rangle$ and then, some time later, when the particles are separated by kilometers, the spin of particle 1 is measured, after choosing some quantization axis. The results are random; but if the spin in that direction turns out to be α, the spin of Fermion 2 *along the same direction* is granted to be β; if it is β, particle 2 must have spin α. The quantization axis can be chosen after the pair is in flight as if the choice of the quantization axis could act on the second particle, which is not true. This is granted irrespective of the distance of the second detector, provided of course that the second quantization axis is parallel to the first.

From a classical viewpoint, this behavior seems to require hidden interaction between the particles, suggesting that Quantum Mechanics fails to give a full description. Worse still, the act of measurement produces an instantaneous collapse of the wave function $|0,0\rangle$, and the interaction (if it existed) should be instantaneous, no matter how large is the distance between the particles at the time of measurement. In principle this distance can be large, and entangled photons emitted by a satellite have recently been found to be entangled at a distance of 1,200 km. According to Relativity, any reasonable interaction should imply a distance-dependent delay, but the collapse does not take time.

This is now called the E.P.R. paradox put forth in 1935 by Einstein–Podolsky–Rosen[2] in a very enlightening paper, which was of enormous benefit to Science for the crisis it produced, although the conclusions of the paper were wrong.

The criticism of EPR is based on two main principles: (1) the principle of realism and (2) the principle of locality.

[2] Einstein et al. (1935).

(1) The principle of realism states that any measurable magnitude has a well-defined value, which is an element of the reality, regardless of the fact that it is measured or not. One can say it in a poetic way: the moon is there, even when nobody looks at it. This is something that does not appear to be questionable.
(2) The principle of locality states that if observers A and B are causally disconnected, in the sense of relativity, no measurement made by A can affect a measurement performed by B. There is no way to do an experiment that produces an instantaneous effect on the Andromeda Galaxy. This principle is made unassailable by an immense amount of the experimental data supporting Relativity.

Quantum Mechanics contradicts both these apparently bomb-proof assumptions. EPR proposed that Quantum Mechanics was providing a description of reality that contained much truth, but not all the truth. The missing truth could be in *hidden variables*. Including such variables, one should render the theory deterministic and complete. The nature of the hidden variables was debated for many years, while the triumph of Quantum Mechanics was complete in all fields of Physics from elementary particles to Astrophysics. Inevitably, this debate on the hidden variables appeared to be in stand-by, as a philosophical, verbose issue.

27.2.1 Bohm's Thought Experiment as a Formulation of the Paradox

In the EPR paper, the paradox was not formulated in terms of spins. In Bohm's reformulation, there are three protagonists who can be at great distance between them. Charlie produces a beam of singlet pairs of Fermions $|\Psi\rangle = \frac{|+-\rangle - |-+\rangle}{\sqrt{2}}$ of spin $\frac{1}{2}$ and sends a member of each pair to Alice and a member to Bob. If Alice finds $\sigma_z^A = +1$, (that is, her particle is in the eigenstate of σ_z^A with eigenvalue 1) the state collapses to $|+-\rangle$ and Bob measures $\sigma_z^B = -1$, while if Alice finds $\sigma_x^A = -1$, the state collapses to $|-+\rangle$ and Bob measures $\sigma_x^B = 1$. Alice's measurement *instantly* changes the state of Bob's particle, whatever the distance. Note that Alice can choose which spin component to measure after the particles have separated. Moreover Relativity says that for some observers Bob's measurement occurs before Alice's; for this observer her decision influences the past!

27.2.2 Plot Twist: Bell's Inequality

In 1964, to everyone's surprise, John Stewart Bell, a Northern Irish physicist, discovered a fundamental theorem[3] showing that the predictions of any hidden-variable

[3] Bell (1964).

theories lead to inequalities that are violated by Quantum Theory. Thus, the issue can be decided experimentally. Let us see how.

27.2.3 Modified Born Thought Experiment for Testing Bell's Inequality

What we need now is an experimental configuration such that the predictions of Quantum Mechanics are at odds with those expected by the EPR criteria. The following thought experiment tests the principle of realism, according to which the values of variables are the same regardless of whether they are measured or not.

Let Alice and Bob (in Bohm's formulation of the Paradox) manage to have parallel axes, perhaps using the stars as reference (the axes used by Charlie do not matter). We shall denote by σ_x^A, σ_z^A the spins of the particles received by Alice and by σ_x^B, σ_z^B those of Bob's particles.

Both Alice and Bob choose at random between measuring the x component of spin or the z component. After recording a large number of results, Alice and Bob compare the two lists. By the results, they compute a quantity M, chosen in such a way that the principle of realism is enough to make sharp predictions, reasoning like EPR.

According to the principle of realism, all the σ values are properties that label the particles and exist independently of measurement, eigenvalues and all that. Since each of them is ± 1, it is clear that both values $(\sigma_z^A + \sigma_x^A)$ and $(\sigma_z^A - \sigma_x^A)$ may be 0 or ± 2. Moreover, if the sum is 0, the difference is ± 2, and if the sum is ± 2, the difference is 0. For this reason, they choose the following quantity:

$$M = (\sigma_z^A + \sigma_x^A)\sigma_z^B + (\sigma_z^A - \sigma_x^A)\sigma_x^B. \tag{27.8}$$

With this choice, each pair yields ± 2, as we have just seen.

Then it is clear that the average over many pairs must be such that

$$|\langle M \rangle| \leq 2. \tag{27.9}$$

This is Bell's inequality in this case. And a result $\ll 2$ seems probable.

Let us compare this with the quantum result, using $\hat{M}$ with the Pauli matrices and $|\psi\rangle = |00\rangle$. Since

$$(\sigma_z^A + \sigma_x^A)(-|1_A - 1_B\rangle) = -|1_A - 1_B\rangle - |-1_A - 1_B\rangle, \tag{27.10}$$

$$(\sigma_z^A + \sigma_x^A)(-|1_A 1_B\rangle) = |-1_A - 1_B\rangle - |1_A - 1_B\rangle, \tag{27.11}$$

and so on, one finds $\langle\psi|M|\psi\rangle = -2$. This is striking, but still marginally in agreement with Bell's inequality (Fig. 27.2).

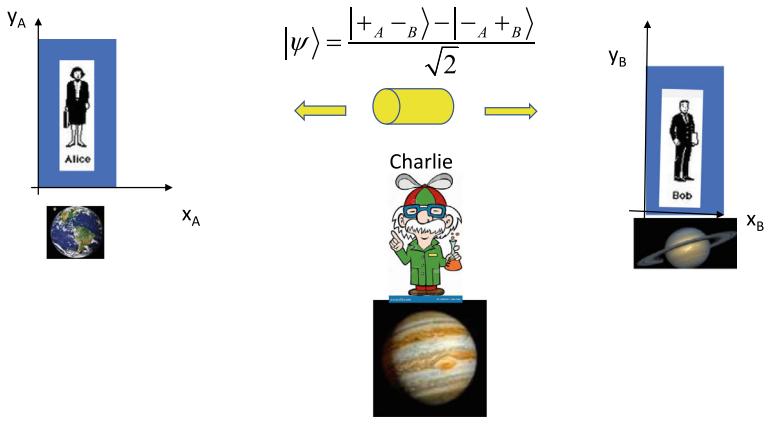


Fig. 27.1 Max Born's thought experiment

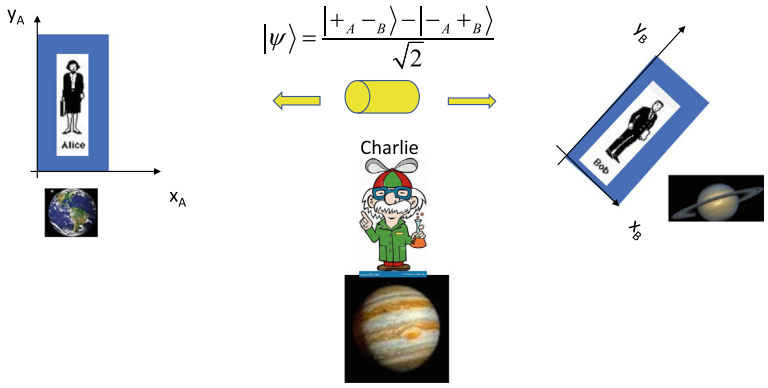


Fig. 27.2 As in Fig. 27.1, except that the reference used by Bob is skewed with respect to that of Alice

Therefore, consider a variant of the Bohm formulation of the EPR paradox in which the reference used by Bob is not parallel to Alice's.

The situation becomes crisp when Alice's reference is parallel to Charle's, but Bob has tilted axes. For the sake of definiteness, assume that, expanding on Alice's basis, the z axis unit vectors are $\tilde{n}_z = \frac{(1,0,1)}{\sqrt{2}}$ and $\tilde{n}_x = \frac{(-1,0,1)}{\sqrt{2}}$. For each pair, both Alice and Bob may decide to measure either the x or the z component of the spin of their particle, and both choose at random and independently, and record the sequence of ± 1 values they obtain (in units of $\frac{\hbar}{2}$, of course). The quantum result is (see next problem)

$$\langle\psi|\hat{M}|\psi\rangle = 2\sqrt{2}. \tag{27.12}$$

This result definitely disagrees with the implications of the principle of realism. If the principle of realism holds, Quantum Mechanics is wrong. Thus, experiment can decide. Experiments have been done and the experimental results support Quantum Mechanics (otherwise the reader would have read the news in the newspaper).

Problem 45 Prove Eq. 27.12.

Solution 45 Since

$$\sigma_x^B = \boldsymbol{\sigma}.n_x^B = \frac{-\sigma_x^A + \sigma_z^A}{\sqrt{2}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ -i & -1 \end{pmatrix},$$

and

$$\sigma_z^B = \begin{pmatrix} 1 & -i \\ i & -1 \end{pmatrix},$$

one contribution is

$$\langle \psi | \sigma_z^A \sigma_z^B | \psi \rangle = \left[\frac{\langle + - | - \langle - + |}{\sqrt{2}} \right] \sigma_z^A \sigma_z^B \left[\frac{| + - \rangle - | - + \rangle}{\sqrt{2}} \right].$$

Now let σ_z^A act on the left spin; since it is diagonal, one obtains

$$\langle \psi | \sigma_z^A \sigma_z^B | \psi \rangle = \frac{1}{2} [\langle - | \sigma_z^B | - \rangle - \langle + | \sigma_z^B | + \rangle].$$

Proceeding in this way, one easily gets the result.

One can interpret the above conclusions as follows. (1) The moon is macroscopic and classical, it is not affected by our observation, so it is there, even when nobody looks at it. However, in the quantum world, the act of measurement assigns quantum numbers that are not predestined before the measurement. (2) A measurement made by A can affect a measurement performed by B, because it causes the collapse of the two-particle wave function. It is entanglement, not a superluminal interaction. Relativity is not in question, anyhow, and no faster than light communication can occur.

27.2.4 The Kochen–Specker Theorem and Quantum Contextuality

The Kochen–Specker Theorem puts further limits (besides those of Bell's inequality) to the possibility to formulate a hidden-variable theory alternative to Quantum Mechanics, such as EPR had assumed, in which all variables have sharp values independent of the instrument used to measure them. The fundamental reason is Quantum

Contextuality. This expression refers to the fact that the measured values of different variables depend on the measurement context, that is, on the choice of variables being measured simultaneously. The basic idea of contextuality was a concern already for the people who started quantum mechanics. Already Bohr stressed "the impossibility of any sharp distinction between the behavior of atomic objects and the interaction with the measuring instruments which serve to define the conditions under which the phenomena appear". By contrast, in Classical Physics one always assumes to be able to make measurements in a clever way, thus avoiding to influence the results.

In other words, the Kochen–Specker Theorem states that non-contextual models, in which the values assigned to the variables are independent of the set of variables that we choose to measure, disagree with quantum theory. This peculiarity, ultimately, depends on the fact that the quantum operators fail to commute, hence the order matters. For more on this topic, see Costantino Budroni et al., arXiv:2102.13036v5 [quant-ph] 9 Jan 2023.

27.3 How the No-Cloning Theorem Saves Coexistence of Quantum Theory with Relativity?

The choice of the quantization axis by Alice instantaneously affects the polarization of Bob's particle and the idea of using entanglement to transmit information instantaneously, circumventing the prohibition that comes from Relativity, is tempting. However it is wrong. Consider an experiment of the EPR type (Sect. 27.2) in which there is a source of pairs of spin $\frac{1}{2}$ particles in the singlet states: suppose that one particle in a pair is sent to Alice and the other to Bob. Alice might decide to send a message to Bob, encoded in a sequence of 0 and 1. To transmit a 0 she measures the x component of the spin, and for every 1 the z component. So Bob receives electrons that are eigenstates of the respective components. The collapse occurs instantly. Now, how could Bob read the message? Measuring the spin once along any axis he cannot come to any conclusion. If Bob were able to make numerous clones of each electron he could find out which is the quantization axis, and read the instant message. If it were possible to clone the state of particles, the instantaneous transmission of information at any distance would be possible and Relativity would be violated. So we may say that the No-cloning Theorem (Chap. 15) saves the coexistence of Quantum Theory with Relativity. No information travels faster than light. On the other hand, since the experimental results support Quantum Mechanics, the principle of realism does not hold and the principle of locality does not work as expected by E.P.R. Note that Relativity is not involved in the proof of the no-cloning theorem.

27.4 Quantum Computer

In 1982, Feynman suggested the possibility of a computer based on the superposition principle and on quantum effects. Instead of the classical bits of Sect. 5.6.2 one could consider two-component spins as quantum bits or qbits. In this context, one writes $|0\rangle \equiv |\uparrow\rangle$ and $|1\rangle \equiv |\downarrow\rangle$. Since one can prepare the qbit in any state of the form of a linear combination $\alpha|0\rangle + \beta|1\rangle$, there are many more ways to prepare the qbit than the classical bit, but when the information is read, it must be read only once, and the result is either $|0\rangle$ or $|1\rangle$. For this reason, the information stored in a qbit is the same as in a classical bit.[4] Qbits can be manipulated by quantum logic gates, analogous to the classical ones.

The quantum analogue of the classical logic gates (Chap. 5) is provided by the evolution of the qbits under the action of some Hamiltonian. The permutation matrices that represent the action of the classical reversible logical gates are replaced by the unitary matrices; for instance, one can use the quantum evolution operator. Therefore, the quantum logical gates are always reversible and their operation costs no energy. For a single qbit, one can use the Pauli matrices as gates. The quantum Toffoli gate is the same as the classical gate and has 3 qbits in input and in output. The real advantages of quantum computing come from the possibility of producing entanglement of qbits. This means that the quantum gate acting on a product of qbits yields a sum of product states.

As an example of a calculation in which a quantum computer should perform better, consider a function $\mathrm{f}_\alpha(x)$ for which the calculation is demanding, but it is known that $\mathrm{f}_\alpha(0)$ and $\mathrm{f}_\alpha(1)$ can be 0 or 1 and that the values depend on the parameter α. Suppose one just needs to know if, for a given α, f_α takes the same value or different values for $x = 0$ and $x = 1$. David Deutsch has shown theoretically that a quantum computer would find the result by one calculation of f. The bonus of quantum computers is the ability to provide global information on the functions to calculate.

Another notable example has important practical applications in cryptography. It is the factorization of large numbers into prime factors. This task requires a number of operations that increase exponentially on a classical computer. The American mathematician P.W. Shor, in 1994, proposed an efficient algorithm for quantum computers that would speed-up the calculation considerably. One could factor an integer N in polylogarithmic time (i.e., the running time is a polynomial in log(N)) while classically the time grows nearly exponentially. So, the quantum computer could lead to an enormous progress; however, there are great difficulties.

A basic problem is due to the fact that the quantum system is never perfectly isolated, and sooner or later is subject to the loss of coherence, or *decoherence* (Sect. 19.6). The idea is the following. Consider a spin in a state $|\psi\rangle = \alpha|\uparrow\rangle + \beta|\downarrow$

[4] This is the physical meaning of the Holevo theorem which puts an upper limit on the classical information that can be extracted by Bob by performing measurements on a quantum state, after Alice has encoded classical information in it.

$\rangle$ with $|\alpha|^2 = |\beta|^2 = \frac{1}{2}$. Our quantum computer could use the information contained in the relative phase of α and β and related to the way the spin evolves in a suitable applied magnetic field. However, if the spin is not perfectly insulated, there is an unknown magnetic field. In other words, part of the Hamiltonian is unknown and describes the interaction with some external system X. The *decoherence* consists in the fact that if X interacts in a different way with $|\uparrow\rangle$ and $|\downarrow\rangle$, and behaves as a detector of the spin state, within some characteristic time, the phase relationship is lost. One can say that the classical behavior emerges naturally when the quantum system is so complex that we cannot know all the Hamiltonian, and then we lose the effects of interference.

27.5 Bell States and Quantum Teleportation

Consider a quantum system (such as a spin $\frac{1}{2}$, for instance) with a two-dimensional Hilbert space; I shall denote the orthogonal states by $|0\rangle, |1\rangle$. A general state of such a qbit is $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, with $|\alpha|^2 + |\beta|^2 = 1$. Now suppose we have two such systems (like an electron spin and a proton spin). A basis for the compound system is provided by the set

$$|0\rangle|0\rangle, |0\rangle|1\rangle, |1\rangle|0\rangle, |1\rangle|1\rangle, \tag{27.13}$$

where the first factors refer to the first subsystem or particle. According to the rules of Quantum Mechanics, one is free to use the entangled basis of Bell states

$$\begin{aligned} \Phi^+ &= \tfrac{|0\rangle|0\rangle+|1\rangle|1\rangle}{\sqrt{2}} \quad & \Phi^- &= \tfrac{|0\rangle|0\rangle-|1\rangle|1\rangle}{\sqrt{2}} \\ \Psi^+ &= \tfrac{|0\rangle|1\rangle+|1\rangle|0\rangle}{\sqrt{2}} \quad & \Psi^- &= \tfrac{|0\rangle|1\rangle-|1\rangle|0\rangle}{\sqrt{2}} \end{aligned} \tag{27.14}$$

and to transform back to the unentangled[5] basis by

$$\begin{aligned} |0\rangle|0\rangle &= \tfrac{\Phi^+ + \Phi^-}{\sqrt{2}} \quad & |0\rangle|1\rangle &= \tfrac{\Psi^+ + \Psi^-}{\sqrt{2}} \\ |1\rangle|0\rangle &= \tfrac{\Psi^+ - \Psi^-}{\sqrt{2}} \quad & |1\rangle|1\rangle &= \tfrac{\Phi^+ - \Phi^-}{\sqrt{2}}. \end{aligned} \tag{27.15}$$

The quantum weirdness already lurks in the equivalence of the entangled basis. We can know that the state of the system is, say, Φ^+, but then we cannot tell the state of individual particles; this is opposite of the classical situation when it is the knowledge of the state of the individual screws and wheels that enables the mechanic to tell the state of a car motor. If we know that the wave function is Φ^+, then all we can grant is that if and only if the first system is in $|0\rangle$, then the second is also in $|0\rangle$.

[5] I choose the example of an electron spin and a proton spin, because the particles must not be identical, if we wish to be able to choose between entangled and unentangled bases.

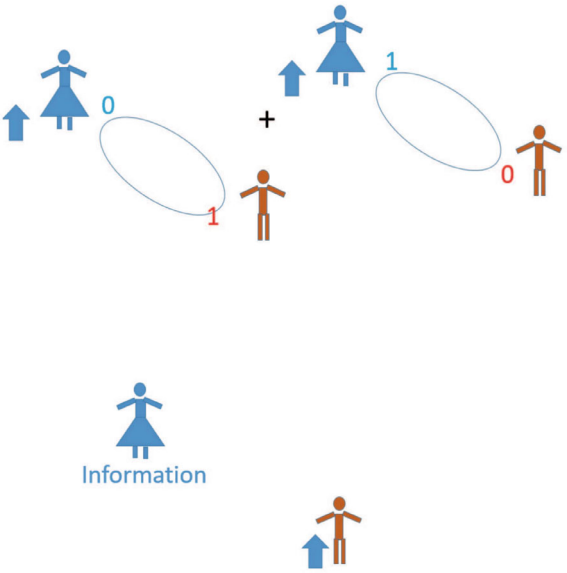


Fig. 27.3 Upper picture: initial situation, with Alice that can handle a qbit and a particle in an entangled pair, and Bob that can act on the other particle. Lower picture: after the measurement by Alice, Bob has the qbit and Alice the information needed by Bob to reproduce the initial qbit state

In 1993 the word teleportation, familiar in science fiction, first appeared in Physical Review Letters[6] in a very often referenced theoretical paper. The thought experiment involves two characters, Alice and Bob (see Fig. 27.3). Alice has a qbit $\psi_A = \alpha|0\rangle_A + \beta|1\rangle_A$; the notation means that only Alice can act on the qbit, which must be transmitted to Bob. Neither Alice nor Bob have any information about α and β, but they share an entangled pair of the Bell Ψ^- type

$$\Psi^- = \frac{|0\rangle_A|1\rangle_B - |1\rangle_A|0\rangle_B}{\sqrt{2}}. \tag{27.16}$$

The three-particle wave function is

$$\Omega_3 = \psi_A \Psi^-, \tag{27.17}$$

where Ψ^- is the Bell state. Expanding $\Omega_3 = (\alpha|0\rangle_A + \beta|1\rangle_A)\frac{|0\rangle_A|1\rangle_B - |1\rangle_A|0\rangle_B}{\sqrt{2}}$, one finds

$$\sqrt{2}\Omega_3 = \alpha|0\rangle_A|0\rangle_A|1\rangle_B + \beta|1\rangle_A|0\rangle_A|1\rangle_B - \alpha|0\rangle_A|1\rangle_A|0\rangle_B - \beta|1\rangle_A|1\rangle_A|0\rangle_B.$$

In terms of Bell states,

$$2\Omega_3 = |\Phi^+\rangle_A(\alpha|1\rangle_B - \beta|0\rangle_B)$$

[6] Bennet et al. (1993).

$$+|\Phi^-\rangle_A(\alpha|1\rangle_B+\beta|0\rangle_B)+|\Psi^+\rangle_A(\beta|1\rangle_B-\alpha|0\rangle_B)+$$
$$|\Psi^-\rangle_A(\beta|1\rangle_B+\alpha|0\rangle_B). \tag{27.18}$$

Now, Alice measures the state of her two particles, thereby performing the instantaneous teleportation according to the following scheme, in which Bob's qbit is rewritten in spinor notation:

Alice's pair	Bob's qbit	$\hat{R}$
$\|\Phi^+\rangle$	$\alpha\|1\rangle-\beta\|0\rangle$	σ_y
$\|\Phi^-\rangle$	$\alpha\|1\rangle+\beta\|0\rangle$	σ_x
$\|\Psi^+\rangle$	$\beta\|1\rangle-\alpha\|0\rangle$	σ_z
$\|\Psi^-\rangle$	$\beta\|1\rangle+\alpha\|0\rangle$	1

If Alice finds Ψ^- (the last case in the table), then Bob has already got ψ_A, which then deserves to be denoted by ψ_B, that is, $\alpha|0\rangle_B+\beta|1\rangle_B$. In the other cases, Bob must apply a restoring operator $\hat{R}$, as shown in the table. He can be sure about the state of his qbit after he receives a call from Alice informing him of the result of her measurement. Therefore, faster than light teleportation is not allowed. Alice does not know the original qbit that she destroyed in the measurement. Not even Bob can know α and β and the No-Cloning theorem (Sect. 15.6) forbids the fabrication of the copies that would be necessary for this investigation. However, the quantum state has been transferred from Alice to Bob. In 2017, Chinese scientists teleported a photon from Tibet to a satellite in orbit, up to 1,400 km above the Earth's surface (see https://www.space.com/37506-quantum-teleportation-record-shattered.html).

27.6 Weak measurements and Quantum Cheshire Cats

Rivers of ink have gushed over the collapse of the wave function as if that were an indispensable element of Quantum Mechanics, but it is also possible to gain information by a gentle non-destructive interaction. Weak measurements are a method for doing so[7]; the information is gained in the quantum regime, and this avoids neatly all the issues that I have mentioned in Chap. 15. The idea is indeed simple. The evolution of the wave function $\psi(t)$ of the system under study is perturbed by a field called by various authors ancilla or pointer, having amplitude $\phi(t)$; quantum measurements on the ancilla provide some valuable information on the system of interest without collapsing its wave function. The overall wave function in the absence of coupling is $\psi(t)\phi(t)$, and once the weak ancilla-system coupling is considered, the change of ϕ can be calculated by perturbation theory in terms of ψ-dependent quantities which are the result of the approach.

[7] Barchielli et al. (1982).

Aharonov and coworkers[8] who pioneered this approach call the ancilla *pointer* and take the interaction of the system and the ancilla to be of the form $H_{int} = g(t)\hat{A}\hat{P}$, where $\hat{A}$ is the system observable that couples to $\hat{P}$, which is the momentum of the pointer. One treats the coupling to lowest non-vanishing order[9] and after some calculations finds that the perturbation is described by the *weak value* of the observable,

$$A^w = \frac{\langle \chi(t_w)|\hat{A}|\psi(t_w)\rangle}{\langle \chi(t_w)|\psi(t_w)\rangle}. \tag{27.19}$$

This is a complex-valued quantity and its physical interpretation is problematic; moreover, the nature of the pointer is not specified. The system in study is assumed to be only gently perturbed, but in order to enhance its characterization, its state is preselected (that is prepared at some initial time) and also postselected (that is, only some final states of the system are considered and the others are discarded). In a work using light in an interferometer as the system under study, Aharonov et al. suggested that a photon could be separated from its polarization in an experiment involving pre- and post-selection [New J. Phys 15, 113015 (2013)]. The photon should be found in an arm of the spectrometer and its polarization on the other. They named the effect "quantum Cheshire Cat", in a reference to the cat that is separated from its grin in the novel Alice's Adventures in Wonderland by Lewis Carroll.

It appears to me that they do not explain clearly what the pointer is and how it interacts in their experiment. Moreover, the separation of an electromagnetic wave from its polarization is a paradoxical result. Instead of having a weak correction of the zeroth-order approximation (photons with their polarization in both arms of the interferometer) they find the photon on one arm and the polarization on the other.[10]

The "quantum Cheshire Cat" interpretation has been questioned[11]; the Authors explained these results as a simple quantum interference, with no paradoxical separation between the quantum particle and its internal degree of freedom. Since this discussion is not settled, and I am not aware of further applications, there is no more space for it in a book about the elements.

[8] Aharonov et al. (1988).

[9] Duprey et al. (2018).

[10] Relativity and Quantum Mechanics have inoculated us from excessive skepticism and accustomed us to strange results; however, we should recall that a weak measurement is based on lowest-order perturbation theory and the results cannot be exhaustive. I should not say that the photon "is" on one arm and the polarization "is" on the other; rather, the probe in the lowest approximation is sensitive to the presence of the photon on one and to its polarization on the other. But this is interesting anyhow.

[11] Correa et al. (2015).

References

Aharonov, Y., Albert, D.Z., Vaidman, L.: Phys. Rev. Lett. **60**, 1351 (1988)
Barchielli, A., Lanz, L., Prosperi, G.M.: A model for the macroscopic description and continual observations in quantum mechanics. Il Nuovo Cimento B. **72**(1), 79–121 (1982)
Bell, J.S.: Physics **1**, 195 (1964)
Bennet, C.H., Brassard, G., Crepeau, C., Josza, R., Peres, A., Wootters, W.K.: Phys. Rev. Lett. **70**, 1895 (1993)
Correa, R., et al.: arXiv:1409.0808v3 [quant-ph] 28 May 2015
Duprey, Q., et al.: arXiv:1703.02959v3 [quant-ph] 26 Jan 2018
Einstein, A., Podolsky, B., Rosen, N.: Phys. Rev. **47**, 777 (1935)

Index

A
Aberration of the light, 131
Aharonov-Bohm effect, 413
Absolute zero, 74, 438
Adiabatic theorem, 268
Angström, 312
Angular momentum, 250
Anharmonic oscillator, 395
Annihilation operator, 293
Anti-bunching, 450
Anyons, 368
Atomic units, 440
Aufbau, 353, 440
Auger effect, 318, 418, 442
Avogadro, 98

B
Balmer, 316
Bardeen, 466
Basis, 242
Bell, 488
Bell, John Stewart, 488
Bell states, 494
Berry , 412
Birkhoff's theorem, 166
Black body, 75, 99, 434
Bogolon, 471
Bohm, David, 488
Bohm–Aharonov effect, 204, 283
Bohr radius, 312, 440
Bohr–van Leeuwen Theorem, 198
Boltzmann, 72, 84, 93, 435, 438
Boltzmann distribution, 95
Born Rule, 263
Born-Oppenheimer approximation, 486
Bose, 352, 433
Bose-Einstein Statistics, 431
Bound states, 231
Breit approximation, 357
Brillouin zone, 321

C
Canonical distribution, 89, 429
Canonical ensemble, 88
Canonical equations, 23
Canonically conjugate variables, 23
Canonical partition function, 96
Canonical temperature, 89
Canonical transformations, 25
Carl Popper, 2
Carnot, 66
Car–Parrinello method, 104
Casimir effect, 299
Cavendish , 153
Cells, 437
Chandrasekar, 445
Chemical potential, 438
Chern numbers, 459
Christoffel, 137
Chronotope, 122
Clapeyron law, 72
Clausius, 66
21cm line, 341
Coherent state, 296
Collapse of the wave function, 212, 330
Commutators, 211, 250
Completeness, 249
Complete set, 247

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

M. Cini, *Elements of Classical and Quantum Physics*, UNITEXT for Physics,
https://doi.org/10.1007/978-3-031-65441-1

Configuration interaction, 397
Continuity equation, 220
Continuum states, 225, 233
Cooper, 466
Cooper pairs, 442
Copenhagen interpretation, 211
Coriolis force, 13
Cosmological constant, 163
Cramer, 234
Creation operator, 293
Current, 220

D
Dark energy, 163
Dark matter, 190
David Ellerman, 207
Davisson, 201
De Broglie, 201, 433
Decoherence, 284, 493
Degenerate stars, 445
Delaunay, 37
Deutsch, 205
Dialogo dei Massimi Sistemi, 107
Diffraction, 199, 202
Dirac, 43, 45, 437
Dirac's δ, 45
Distinguishable particles, 93
Distribution function, 82
Doppler effect, 129
Doppler–Fizeau effect, 129
Double slit experiment, 203
Drude, 453
Drude theory, 60
Dulong, 99, 437

E
Ehrenfest, 151, 158
Ehrenfest paradox, 152
Ehrenfest Theorem, 218
Eigenvalue equation, 209
Eigenvector, 247
Einstein, 99, 107, 109, 201, 283, 431, 437, 487
Electromagnetic waves, 199
Electron spin resonance, 338
Energia di Fermi, 440
Energy–Momentum Tensor, 136
Energy operator, 216
Ensemble, 82
Entropy, 68, 73, 77, 85, 96
E.P.R. paradox, 487
Equilibrium, 65
Equipartition theorem, 98
Ergodic problem, 82
Euler, 9
Euler–Lagrange equations, 9, 33
Evolution operator, 217
Exchange, 356, 441
Expectation value, 213, 247

F
Fano, 366
Fermat, 20
Fermi, 352, 438, 445
FermiDirac, 437
Fermi–Dirac Statistics, 437
Fermi energy, 440
Fermi function, 438
Fermi gas, 438
Fermi level, 438
Fermion annihilation operator, 362
Fermion creation operator, 362
Fermions, 429
Feynman, 278, 394, 493
Field Lagrangian, 136
Field operators, 364
First Principle of Thermodynamics, 65
Flux quantum, 403
Fourier, 200
Fractional quantum Hall effect, 369
Free energy, 69
Fritz Zwicky, 190
Fröhlich , 463
Fundamental commutation rule, 211

G
Galileo, 5, 107
Gauge invariance, 52
Gauss, 137
Geim, Andre, 322
Generalized momenta, 12
General Relativity, 151
Generating function, 26
George Green, 53
Gerlach, 328
Gibbs, 80, 430
Gibbs free energy, 69
Gibbs paradox, 86
Goldstone, Jeffrey, 460
Gram–Schmidt orthogonalization, 242
Grand-Canonical ensemble, 101
Green's function, 52
Grimaldi, Francesco Maria, 199

Group velocity, 227
Gustav Kirchhoff, 75
Gyromagnetic factor, 327

H
Hall, 60
Hamilton, 22
Hamiltonian, 22
Hanbury Brown and Twiss effect, 450
Hanbury Brown and Twiss experiment, 60
Hartree–Fock, 395
Hartree–Fock method, 319
Hawking, 448
He, 357
He 3, 446
He 4, 433
Heavyside, 43
Heisenberg, 203, 222
Herglotz, 422
Hermite, 292
Hermitian, 247
Hilbert space, 244, 245
Holevo theorem, 493
Holonomic constraint, 7
Hubble, 131, 184
Hyperdifferential operator, 256
Hypersphere, 49, 86
Hypersurface, 85

I
Independent variables, 12
Indistinguishable particles, 429
Inertial frames of reference, 108
Integrable systems, 39, 41
Internal energy, 65, 96
Interval, 122
Involution, 39

J
Jacobi, 29, 36
Jacobi identity, 36
Jeans, 99
John Wheeler, 206
Josephson effect, 237
Joule, 65

K
Kelvin, 66
Kepler, 312
Kepler problem, 16, 38
Kerr solution , 179
Kramers degeneracy, 349

L
Lagrange, 8, 95, 438
Laguerre, 317
Lamb shift, 299
Landau, 305
Landauer principle, 104
Laplace, 41
LCAO method, 395
Legendre, 21, 69
Legendre transformation, 21
Levi Civita, 137
Lienard, 56
Linear, 247
Liouville, 82
Logic gates, 104
Lorentz force, 32
Lorentz transformation, 116
Löwdin symmetric orthogonalization, 242

M
Madey,John, 436
Magnetic Moment, 327
Many-body Hamiltonian, 365
Many-worlds interpretation, 286
Matrices, 247
Maupertuis, 20
Maxwell, 51, 71, 80, 199
Maxwell's equations, 51
Meitner, 442
Meitner-Auger effect, 442
Metals, 439, 441
Microcanonical ensemble, 437
Microstates, 85
Minimal coupling, 34
Modello LCAO, 401
Molecular Aharonov–Bohm effect , 413
Moment of inertia, 15
Momentum operator, 209
Monte Carlo methods, 102
MOSFET, 238
Mössbauer effect, 132, 177
Mossbauer effect Mössbauer effect, 132

N
Neutrino oscillations, 402
Newton, 5, 199
No Cloning Theorem, 277
Noether's theorem, 39

Normalization, 225
Novoselov, Andre, 322
Nuclear spin resonance, 338

O
Observable, 210, 241
Onsager, 463
Oscillator, 289

P
Pancharatnam , 409
Parity, 309
Partition function, 91
Pauli, 332
Pauli equation, 337
Penzias,Arno , 188
Perfect gas, 72, 93
Periodic boundary conditions, 226
Permanent, 362
Petit, 99, 437
Phase, 409
Phase space, 81, 89, 430, 431
Phase velocity, 226
Photoelectric effect, 201
Photon, 434
Photon bunching, 430
Piezoelectric effect, 239
Pisa, 5
Planck, 66, 77, 197
Planck length, 300
Planck mass, 300
Planck's law, 435
Plasmon, 367
Podolsky, 487
Point transformation, 12, 25
Poisson, 35, 250
Pontecorvo, Maki, Nagakawa, Sakada matrix, 402
Positive definite operator, 229
Pressure, 97
Principal quantum number, 316
Principle of equivalence, 151
Probability amplitude, 209

Q
Quantum computer, 284
Quantum numbers, 245
Quantum Runge–Lenz vector, 312
Quantum tomography, 214
Quintessence, 190

R
Rabi model, 346
Rayleigh, 99
Reduced mass, 24
Reversible transformation, 66
Ricci, 144
Ricci Curbastro, 137
Riemann, 137
Rigid rotor, 7
Risonances, 417
Roger Penrose, 182, 446
Rosen, 487
Rotating Wave Approximation, 347
Rotor, 7, 15, 250
Runge–Lenz vector, 39
Rydberg, 441

S
Schrödinger, 203, 222, 241, 394
Schrödinger equation, 265
Schrödinger's cat, 284
Schrieffer, 466
Schwarz, 244
Schwarzschild radius, 166
Second principle, 66
Selection rules, 384
Self-energy, 421
Shapiro, 175
Shapiro effect, 238
Simultaneity, 111
Single electron transistor, 238
Singlet, 340, 487
Solid, 99
Sommerfeld, 439, 441
μ space, 93
Γ space, 85
Specific heat, 72
Spectrum, 99
Spherical coordinates, 306
Spherical harmonics, 308
Spin, 327
Spin-orbit interaction, 342
Spontaneously broken symmetry, 467
Standard deviation, 213
Stark–Lo Surdo Effect, 380
Stationary State Equation, 218
Stefan, 435
Stefan–Boltzmann Law, 77
Stern, 328
Stirling's formula, 87, 94
Stress–Energy Tensor, 136
Superconducting Quantum Interference Devices (SQUID), 238

Superconductors, 237
Superposition principle, 208
Superfluid, 433

T
Tensor product, 245
Tensors, 125
Thermodynamic cycle, 66
Third principle of thermodynamics, 86
Thomas Young, 199
Thouless, David, 458
Toffoli gate, 105
Topological quantum computer, 369
Townes, Charles H., 436
Transition probability, 382
Triplet, 340
Tunnel effect, 235
Tunnel effect microscope, 238
Twin Paradox, 114
Two-dimensional systems, 368

U
Ultracold Atoms, 434
Ultraviolet catastrophe, 99, 201
Unitary operators, 255
Unruh effect, 448

V
Variational principle, 19, 22
Vera Rubin, 189
Virtual orbitals, 397
Virtual paths, 17
Vladimir Aleksandrovich Fock, 165
Von Klitzing, Klaus, 454

W
Wave function, 208
Wave packet, 227
Wave-particle dualism, 201
Wiechert, 56
Wien's Law, 75
Wigner–Seitz radius, 440
Wilson, Robert, 188

Z
Zel'dovich , 190
Zero point energy, 292
Zero sound, 438
Zero-th Principle of Thermodynamics, 65